0	NLINE SEARCH REQUES	GAL I T FORM * * * * * * * * * * * * * * * * * * *	****	
**************************************	* * * * * * * * * * * * * * * * * * *	17 223263		
UULII —————	PHONE 308-1793	DATE 4/10/93		
Please give a detailed statem matter to be searched. Define	ent of requirements. Describe any terms that may have spe s, if known.		ole the subject les or relevant	
You may include a copy of the	broadest and or relevant claim	s). sL7 —		
Please	broadest and or relevant claims	V 22		
/ / -				
			J. See J.	
d .				
. *				
· ••				100 mg
•				ر ا
		•	·	
				1
	and the second s	The state of the s		
		and the second s	The state of the s	
	·			
		· · · · · · · · · · · · · · · · · · ·		1 m
		•		
				َ الْوَبَ
*****	* * * * * * * * * * * * * * * * * * *	*****	******	 .
	STAFF USE ONL	SYSTEMS	4-1	
COMPLETED 4	XUICY	CAS ON	UESTEL	
SEARCHER	TIME	DIALOG SDC ~ /OTHER	τ_c	

一名的 医乳化二甲 计记录 医二种

••:

Tue Apr 11 13:27:35 1995

iGenetics

Fast Pairwise Comparison of Sequences

Results file sqlasq.res made by on Tue 11 Apr 95 9:45:54-PDT.

Query sequence being compared:US-08-223-263-1 (1-353)
Number of sequences searched: 50375
Number of scores above cutoff: 4542

Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : A-GeneSeq 17, all entries

100000-N -U50000-M -B -E -R -

O - F10000- - S - -

500-

Listing for Mary Hale

SCORE 01 STDEV -1 072 ω₁= 4 515 17| 20 7

22 –

PARAMETERS

	Initial scores to save Optimized scores to save	Similarity matrix Un Mismatch penalty Gap penalty Gap size penalty Cutoff score Randomization group
SEAL	45 0	Unitary 1.00 0.05 2
SEARCH STATISTICS	Alignments to save Display context	K-tuple Joining penalty Window size
	15 10	32 32

Cut-off raised to 3. Cut-off raised to 4. Cut-off raised to 5. Cut-off raised to 7. Cut-off raised to 8. Cut-off raised to 9.	Number of residues: Number of sequences searched: Number of scores above cutoff:	Times:	Scores:	
	searched: ove cutoff:	CPU 00:00:50.06	Mean 4	OPAN
	6065180 50375 4542		Median 5	OFFICE STRITTSTICS
		Total Elapsed 00:00:51.00	Standard Deviation 2.39	

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

2. R29165	1. R29164	Sequence Name
165	64	ame
PRP3 (from genomic clone PRP3 151 20	**** 7 standard deviations above mean **** PRP3. 121 22 PRP3. 121 22 **** 6 standard deviations above mean ****	Init Opt. Description Length Score Score Sig. Frame
20	n **** 22 ****	re Sc
34	30	ore
6.71	7.54	Sig. F
0	0	rame

Page

45. R37486	•	•	•	•	•	•	•	•	•	•	•	•	32. R54979	•		30. R04032																14. R14838					•	83	7. R45155		. R2022	. R2815	4. P50343	. p5030
y acyl reductase	of alkaline phos	hematopoietic	inant hematopoietic	enocorticot	and heavy chair	Non	Cysteine-added variant of ery	Human recombinant erythropoie	binant	Sequence of human erythropoie	hemato	Sequence of plant acyl carrie	ica seed acyl	Light (kappa) chain variable	ns a	gth T4 encoded by pl	encoded by the 2nd	ia rickettšii p120	encoded by I	portion of E	qondii P66 anti	protein 1	polysacc	growth	growth hormone.	growth	growth hormone.	. of DNA of	IIV qp120 immunoc	n Cytomegalovirus	~	deduced from human	in a low affinity	receptor from	nal mucin deduced	87 intestinal mucin	associated with	acillus caldotenax DNA p	thermophilus DNA poly	**** 5 standard deviations ab	•	chitinase	coded by clone	Monkey erythropoietin.
524	409	322	303	266	241	205	167	167	167	166	166	134	134	128	bove mean	2037	1594	1267	859	516	428	392	311	204	200	188	188	156	106	1048	635	405	360	355	119	119	900	877	876	940	350	439	193	192
15	J. 1	15	15	15	5	15	15	15	15	15	15	15	15	15	***	16	16	16	16	16	16	16	16	16	16	16	16	16	16	17	17	17	17	17	17	17	18	18	18	****	19	20	20	20
48	J.	64	63	42	31	26	38	38	38	38	38	28	28	27		61	59	66	72	53	44	45	48	43	41	39	39	36	25	64	33	71	39	39	30	30	64	60	60		60	40	44	43
4.61				•	٠			4.61	٠	٠	4.61					•				5.03										٠.	<u>.</u>	5.45	<u>.</u>	٠.	.4	٠.	٠	5.87			'n	٠.,	6.71	-1
00	> -	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0

US-08-223-263-1 (1-353) R29164 PRP3.

R29164 standard; Protein; 121 AA. R29164; R29164; 21-APR-1993 (first entry) PRP3.

Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12; pistil; flower; sterile; plant; hydroxyproline; glycoprotein; gum. Nicotiana alata. Location/Qualifiers Region 23..24
/note= "Ser-Pro repeat unit"

Monkey erythropoietin Tue Apr 11 13:27:35 1995 192 20 43 6.71 ω 0

Listing for Mary Hale

Tue Apr 11 13:27:35 1995

SQ	888	888	88	88	888	38	88	Sq	T T	ΡŢ	DR PR	DR	ΡA	PR	PF Od	PN	H 1	# F3	H	H H	FΤ	H H	FT	H F	1 H	H 1	F	F F	H H	1 1	F F	F F	1 H	H 1	7 F	d 4
	where emulsification, thickening and stabilisation are required. Sequence 121 AA;	erted by hydroxylation and glycosylation into hydrox opportains. These could by used as effective substitu-	. to enable longer flower life- -sterile plants. The PRPs may	Such elements may be used to provide pistil-specific expression of	PRP4 genes are given in 031794-98. The DNA sequences may be used to	repeats.	proline (37%) and serine (17%). The sequence comprises 6 Ser-(Pro)4,	3 1-DBD3 6 is rich	pistil specific expression, to give e.g.	to express useful proteins and identify r	no opposite proline with protein	WPI; 92-415710/50.	(UYME) UNIV MELBOURNE.	· 🚡	26-NOV-1992. 15-MAY-1992: AU0222.		/note= "Ser-Pro repeat unit"	"Ser-Pro repe	9192	8284	"Ser-(Pro)	/note= "Ser-Pro repeat unit" Region 7781	7576	/note= "Ser-(Pro)4 repeat unit"	"Ser-Pro	Region 6869	6367	Region 6162 /note= "Ser-Pro repeat unit"	"Ser-	"Ser-(Pro)4	/note= "Ser-(Pro)4 repeat unit" Region 5155	4448	3538	/note= "Ser-Pro repeat unit"	"Ser-Pro repe	3

Tue Apr 11.13:27:35 1995

Page Ġ

Gaps Residue Initial PSPSPPPPSPSPSPP-PPSPPADDMAPSPSP-AAAPTPPDMLIRPSARLAGNMVISCIH
60 70 80 90 100 110 120 230 X 240 250 260 RAKIPGLINQTSRSIDQIPGYLNRIHELLNGTRGLFP-300 310 320 330 340 350 XSPSPTHPPTGQYTLFPLPTPTVQLHPLLPDPSAFTPTPTSPLLNTSYTHSQNLSQEG Identity Score GTESL--PDATIOR-FKFKWPFFGKSPKNSPKSSPSRSPPPKREQPSPPPPVKSPPPPSPPP X 10 20 30 40 50 11 11 () 22 28% 11 Matches = 35 Conservative Substitutions Optimized Score lį --GPSRRTLGAPDISSGTSDTGSLPPNLQPG 35 270 Mismatches Significance = 280 .54 78 0 290

US-08-223-263-1 (1-353) R29165 PRP3 (from genomic clone PRP3g12). R29165 standard; Protein; R29165;

151 AA

21-APR-1993 (first entry)
21-APR-1993 (from genomic clone PRP3g12).
PRP3 (from genomic clone PRP3g12) PRP3; PRP4; PRP3g12; pistil;
Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12; pistil;
flower; sterile; plant; hydroxyproline; glycoprotein; gum. Region /note= Region Region Peptide Nicotiana alata. Region Region /note= Region /note= /label= sig_peptide note= note= note= "Ser-Pro *Ser-Pro "Ser-(Pro)4 repeat unit" 81..85 "Ser-Pro "Ser-Pro "Ser-(Pro) "Ser-(Pro)3 repeat unit" repeat unit* 88..92 repeat unit *61..62 repeat unit 7 57..58 b)4 repeat unit 86..87 repeat unit* 65..68 1..23

Region /note= Region Region Region note= note= note= "Ser-(Pro)4 repeat unit" 98..99 "Ser-(Pro)4 repeat unit" 105..106 "Ser-(Pro)4 repeat unit" 93..97 "Ser-Pro repeat unit 100..104 repeat unit*
107..111

Region

Region "Ser-Pro repeat u 123..124 "Ser-(Pro)2 repeat unit" 121..122 unit*

Region Region

"Ser-(Pro)4 repeat unit"

112..114

"Ser-Pro

1 10

Listing for Mary Hale

Tue Apr 11 13:27:35 1995

6

```
Initial Score
Residue Identity
Gaps
                                                                                                                                                                                                                 15-MAY-1992; AU0222.
15-MAY-1991; AU-006159.
(UYME) UNIV MELBOURNE.
Chen C, Clarke AE;
WPI; 92-415710/50.
RTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFP-----G
                                                                                                                                                                                                                                                                                          PRP4 genes are given in Q31794-98. The DNA sequences may be used to identify regulatory elements responsible for pistil-specific expression. Such elements may be used to provide pistil-specific expression of foreign or heterologous genes, e.g. to enable longer flower life or in the creation of male- or female-sterile plants. The PRPs may be converted by hydroxylation and glycosylation into hydroxyproline-rich glycoproteins. These could by used as effective substitutes for gums, e.g. guar gum and gum arabic, in foods and for non-food applications, where emulsification, thickening and stabilisation are required.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Claim 2; Pag 47 + Fig 5; 57pp; English.

The sequence below is derived from the genomic clone PRP3g12.

The polypeptide of 15,645 D is rich in proline (35%) and serine (16%). The sequence comprises 6 Ser-(Pro)4, 1 Ser-(Pro)3, 1 Ser-(Pro)2 and 8 Ser-Pro repeats.

1 Ser-(Pro)2 and 8 Ser-Pro repeats.

The sequences of the PRP1, PRP2, PRP3 (cDNA and genomic DNA), PRP4 genes are given in Q31794-98. The DNA series may be use the pressure of the pressure of the PRP4 genes are given in Q31794-98. The DNA series may be use the pressure of the pr
                                                                                                                                                                                                          Sequence 151 AA;
10 A; 5 R; 4 N;
3 I; 12 L; 8 K;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Isolated DNA sequence encoding proline rich proteins of Nicotiana alata - used to express useful proteins and identify regulatory elements for pistil specific expression, to give e.g. longer
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           N-PSDB; Q31797.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     flower life
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               WO9220713-A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            26-NOV-1992.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        'note= "Ser-Pro repeat unit"
                                                                                        20
24%
11
                                                                                                                                                                                                              м;
50
                                                                                        Optimized Score = 34
Matches = 38
Conservative Substitutions
                                                                                                                                                                                                              শ্ল
                                                                                                                                                                                                       45
                                                                                                                                                                                                              ρÇ;
                                                                                                                                                                                                          3 Q;
26 S;
                                                                                                                                                                                                                 ıω
                                                                                                                        34
38
                                                                                                                                                                                                                 ΗĦ
                                                                                                                  Significance
Mismatches
                                                                                                                                                                                                                 - 0
                                                                                                                                                                                                                 Σ,
                                                                                                                                                                                                                 0 5
                                                                                                                                                                                                              <u>۲</u>.6
                                                                                                                                                                                                                 σN
                                                                                                                                                     11
                                                                                               11 (1
                                                                                                                                                                                                              <. E
                                                                                                                                                     6
                                                                                        104
104
```

PDMLLRPSARLAGNMVISGLH 140 150 TSPLLNTSYTHSQNLSQEG PSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPS-APTPTP 350

3. US-08-223-263-1 (1-353) P50301 Monkey erythropoietin.

AC AC P50301 standard; protein; 192 AA P50301;

Com

01-JAN-1980

Tue Apr 1 13:27:35 1995

Page

Residue Initial 11-DEC-1984; U02021.
13-DEC-1983; US-561024.
21-FEB-1984; US-582185.
28-SEP-1984; US-655841.
30-NOV-1984; US-675298.
(KIRI-) KIRIN-AMGEN INC.
WPI; 85-159229/26.
N-PSDB; N50348. Sequence 192 20 A; 13 R; 6 6 I; 30 L; 7 Monkey. W08502610-A. disorder; Erythropoietin; red blood cell; erythrocyte; anaemia; blood; 20-JUN-1985. Monkey erythropoietin. Identity Score ds. 192 AA; (first entry) X ; Optimized Matches K D 50 'E D 11 Score ΑĊ. 7 14 H H ŝö 9 13 54 3 HE 0 4 Mismatches Significance ΣZ 12 ۲. G 13 -tt < H

P50343 standard; Protein; 193 A

Gaps SEEDEL US-08-223-263-1 (1-353) P50343 EPO encoded by clone lambda-HEPOFL13. VNFYAWKRMEVGQQAVEVWQGLALLSEAV--LRGQAVLANSSQPFEPLQLHMDKAISGLRSITTLLRALGAQ MGVHECPAWLWILLISIVSIPIGIPVPGAPPRLICDSRVIERYLLEAKEAENVTMGCSESCSINENITVPDTK X 10 20 30 40 50 60 70 EAISLPDAASAAPLRTITADTFCKLFRVYSNFLRGK--150 160 170 FSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQL--MELTELLLVVMLLL-TARLTLSSP---APP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLFAVD 130 140 150 160 170 180 - LOSILGTQHELRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTS New polypeptide having properties of erythropoletin — is prepd. by cultivation of transformed eucaryotic or procaryotic host Disclosure; Page 38; 113pp; English.

Monkey erythropoletin encoded by this sequence is essential for red blood cell formation and is used for the diagnosis and treatment of blood disorders such as anaemia. Large amounts of EPO may be obtained using recombinant DNA techniques in contrast to small amounts obtained from plasma and urine. See also N50345-47, N50349-50 and P50298-P50300. Erythropoietin. Homo sapiens. EPO encoded by clone lambda-HEPOFL13 0-MAR-1992 250343; 90 11 11 11 (first entry) 20 27**%** 27 20 90 ** Conservative Substitutions 100 10.46 --LKLYTG-EACRRGDR 180 190 X 50 SG---QVRLL--LGA-120 $\Pi = \Pi$ 60 6.71 117 0

Listing for Mary Hale

Tue Apr 11 13:27:35 1995

2 1

Page

Ø

Initial Residue Gaps VNFYAWKRIEVAQQAVEVWQGLALLSEAV--LRGQALLVNSSQPWEPLQLHVDKAVSGLRS-LTTLLRPLGA 80 90 100 110 120 130 140 QKEAISPPDAASAAPLRTITADTFRKLFRLYSNF--LRGKLKL--150 160 170 180 IGVHECPASIMILISIPIGLPVRGAPPRLICDSRVLLERYLLEAKQAENITTGCAEHCSINENIIVPDTK X 10 20 30 40 50 60 70 RTTAHKDPNA-- ${ t FSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQG-$ Sequence 20 A; 14 8 I; 33 cells - comprises prod. gene and accessory DNA for enhanced expression of heterologous protein by the cells. Disclosure, Fig 13; 62pp; Danish. See also US4740461 88.04.26 (8819) (first major country equivalent). The sequence is encoded by clone lambda-HEPOFLI3 which contains the gene encoding erythropoietin obtd. from a human foetal DNA library. The gene can be expressed using a novel vector system dislosed in the specification.

See also P50342. Vector system for introducing heterologous DNA into eukaryotic cells — comprises prod. gene and accessory DNA for enhanced N-PSDB; N50443 DK8406107-A WPI; 85-318061/51. (GENE-) GENETICS INST. Kaufman RJ. 22-JAN-1985; 22-JAN-1985; 21-DEC-1984; 21-DEC-1984; 04-DEC-1984; 27-DEC-1983; 27-DEC-1983; 27-DEC-1984; /label= .6-AUG-1985 'label= mature_EPO MELTELLLVVMLLLTARLTL-SSPAPP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVD Identity signal_sequence 27..193 1 R; 6 193 H H H US-565627. US-566057. US-677813. NO-005186. ZA-010034. US-693258. US-688622. 281862 ZZ. AA; 26**%** 26**%** 25 Location/Qualifiers 1..27 0 0 20 Optimized Matches Conservative Substitutions K O 40 ㅋ♡ 4 12 Score 40 8 14 11 11 S S 40 10 53 53 ΗE --YTGEASRTGDR 190 X Significance Mismatches 40 ₹,2 411 ₹.6 3 12 11 II II < #

6 5.71 124 0

US-08-223-263-1 (1-353) R28150 Sugar beet chitinase 1.

H R28150 standard; Protein; 439 AA



Key

Peptide

/label= leader

Sugar beet chitinase 1.
SBC-1; fungicide; anti-fungal agent;

extensine.

17-MAR-1993

(first entry)

Beta vulgaris cv monova.

Location/Qualifiers

Listing for Mary Hale

Tue Apr 11

13:27:35

1995

10

1

1

```
Residue
                                                                                                                                                                                                                             Initial
                                                                                                                                                                                                                                                180 X 190 200 210 CVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASAR-
                                                                                      DQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPT
                                                                                                                                                                                                                                                                                   germination and growth of chitin-containing fungi and is used to produce genetically transformed plants Claim 5; Page 164-168; 254pp; English.

A clone encoding the chitinase 1 gene was isolated from a sugar beet EMBL3 genomic library. The sequence encodes a protein having 439 amino acid residues. Transgenic plants having increased resistance to nematodes and chitin-containing plant pathogens, partic. fungi, can be produced using genetic constructs containing the chitinase 1 gene. The protein itself can be used in fungicidal
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Domain
/label= hevein_domain
47.178
                                                                                                                                                                                                                                             Sequence
15 A; 33
19 I; 26
                                                                                                                                                                                                                                                                                                                                                                                            Berglund I, Bojsen K, Mikkelsen JD, WPI, 92-366261/44.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 extensines"
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         glycosylated hydroxyprolines,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     chitinase 1 protein to the cell wall after modification of the prolines to
                                                                                                                                                                                                                                                                                                                                                                           Sugar beet chitinase 4 and
                                                                                                                                                                                                                                                                                                                                                                                    N-PSDB;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                the vacuole
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Region
                                                                                                                                                                                                                                                                           compositions.
                                                                                                                                                                                                                                                                                                                                                                                                                            08-APR-1991;
                                                                                                                                                                                                                                                                                                                                                                                                                                    07-APR-1992; DK0108
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   /label= functional domain Region 417..439
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      /label= proline_rich
/note= "possibly involved in anchoring
                                                                                                                                                                                                                                                                                                                                                                                                                (DANI-) DANISCO AS.
                                                                                                                                                                                                                                                                                                                                                                                                                                                        ₩09217591-A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                         /note= "probably
                                                                                                                                                                                                                Identity
                                                                                                                                                                                                                           Score
                                                                                                                                                                                                                                              33 R;
26 L;
                                                                                                                                                                                                                                                                                                                                                                                    Q29965
                                                                                                                                                MKIKTS-PSFLLGLICLALVLLLGEGVQCGRQCNTTDTNCLSGCSVGRPSRP X 10 20 30 40 50
                                                                                                                                                                                                                                              439 AA;
16 N;
; 8 K;
                                                                                                                                                                                                           0 11 0
                                                                                                                                                                                                                                                                                                                                                                                                                            DK-000616.
                                                                                                                            260
                                                                                                                                                                                                         20
25%
19
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         179..416
                                                                                                                                                                                                                                                                                                                                                                                                                                                                          directs
                                                                                                                                                                                                                                              <sup>3</sup> 13
                                                                                                                                                                                                                                               K O
                                                                                                                                                                                                         Conservative Substitutions
                                                                                                                                                                                                                            Optimized
                                                                                                                                                                                                                                                19
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           the
                                                                                                                             270
                                                                                                                                                                                                                                                                                                                                                                 chitin-containing fungi and it
                                                                                                                                                                                                                                               Ŧ,B
                    Physical Property of
                                                                                                                                                                                                                                                                                                                                                                                                                                                                          protein
                                                                                                                                                                                                                                              12 C;
109P;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           as in
                                                                                                                                                                                                                             Score
                                                                                                                            280
                                                                                                                                                                                                                                                                                                                                                                                                         Nielsen
                                                                                                                                                                                                                                              14
23
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           ç
                                                                                                                                                                                                                    n ti
                                                                                        90
                                                                                                                                                                                                                                               Sign
                                                                                        Link
                                                                                                                                                                                                                                                21
39
                                                                                                                                                                       240 230 240
TTGSGLLKWQQGFRAKIPGLLNQTSRSL
                                                                                                                                                                                                                                                                                                                                                                                                         <del>汉</del>
                                                                                                                                                                                                                   40
                                                                                                                                                                                                                                               ΗE
Mir.
                                                                                                                            290
                                                                                                                                                                                                                  Significance
Mismatches
                                                                                                                                                                                                                                                90
                                                                                                                                                                                                                                               ₹2,
                                                                                                                                                                                                                                               31
12
                                                                                                                            300
                                                                                                                                                                                                                                              મ્ છ
                                                                                                                                                                                                                                               7
10
                                                                                                                                                                                                                            II
                                                                                                                                                                                                                                               < #;
                                                                                                                                                                                                                             6.71
```

Initial Residue

Score Identity

0 0 0

19 20% 66

Optimized Matches

Score

60 79

Significance Mismatches

11 11

6.29 233 0

6

Conservative Substitutions

Sequence 44 A; 19 9 I; 53

Ľ,R

7 Z

9 13

3 0

22

Ŧ, B. 13

79.0

9 18 S O

11 12

ΗM

13

Σ.

16 21

₹.6 12 21

<....

350 6

Claim 18; Fig 11A; 58pp; English.

The protein sequence was deduced by screening adult mouse brain cDNA libraries with a 1.4kb GDF-1 sequence isolated from a mouse embryo CDNA library (see Q20688). The DNA sequence of 2.7kb obtd. comprised two ORFs, one of which was highly homologous to the 1.4kb probe.

(Q20688). Upstream of the GDF-1 coding region is a 2nd ORF, encoding mUGG-1 (Upstream Of GDF-1) which terminates upstream of the GDF-1 ATG codon. Due to the presence of multiple stop codons in the region between mUGG-1 and mGDF-1, at least 4 frameshifts would be required to translate the two ORFs as a single protein. Hydropathicity analysis of the amino acid sequence of mUGG-1 revealed multiple clusters of hydrophobic residues, indicative of membrane spanning domains. mUGG-1 does not contain an obvious N-terminal signal US-08-223-263-1 (1-353) R20229 UOG-1. TEREPPEPEPEPEPEPEPPEPPEPPEPEPTPSPEPEPTPEPEPTPT 130 140 150 160 X 170 320 330 340 350 x LPTPVVQLHPLLPDPSAPTPTPTSPLLNTSYTHS--QNLSQEG WPI; GDF-1 DNA segment encoding a mammalian GDF-1 protein -tumours, birth defects and genetic diseases. N-PSDB; Q20689 14-JUN-1991; U04096. 15-JUN-1990; US-538372. 16-NOV-1990; US-614452. 09-JAN-1992. R20229 standard; Protein; (CARN-) CARNEGIE INST 06-MAY-1992 See also sequence. 92-041563/05. .; mouse; R20228, 30 and R22375. (first murine; hydrophobic clusters entry) WASH 350 B for diagnosis of



VDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPP-

MELTELLLVVMLLLTARLTLSSPAP-PACDL--RVLSKL-----LRDSHVLHSRLSQCPEVHPLPTPVLLPA

40

50

8

```
200 210
TSGLIETNFTASARTTGSGLL-
                                                                                                                                                                                                                                                                                                                                                                                                                                                      VFYDWRSGMAVPWDIAVAYLLQGSFYCHSIYATVYMDSWRKDS-
140 150 160 170
TLEAQTAKPCKAEKPLRNGLVKDKLF
330 340 350
                                                                        DPSAPTPTP---TSPLLNTSYTHSQNLSQEG
                                                                                                                                                                           LYWFP----LKVLYATCHCSLQSVPDIFYYFFFNILLLLLMVMNIYWFLYIVAFAAKVLTGQMRELEDLREYD
                                                                                                                                                                                                                       260 270 280 290 300 310 320 RGLFPGPSRRTLGAPDISSGTS--DTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPTLPTPVVQLHPL--LP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               VLCALGWTALRWAATTHIFRPLAKRCRLQPRDAARLPESAWKLLFYLACWSYCAYLLLG--TSYPFFHDPPS
                                                                                                     330
                                                                                                                                                                                                                                                                                                     -LLIASSYA-FRYHNVGLLVFFLHDVSDVQLEFTKLNIYFKARGGAYHRLHGLVANLGCLSFCFCWFWFR 190 200 210 220 230 240 250
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             140 150 160 170 180 190
-QGRTTAHKDPNAIFL--SF-QHLLRGKVRFIMLVGGSTLCVRRAPPTTAVPSRTSIVLTLNELPNR
                                                                                                     340
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          9
                                                                                                                                                                                                                                                                                                                                                                                                     220
                                                                                                                                                                                                                                                                                                                                                                           -KWQQGF----RAKIPGLLNQTSRSLDQIPGYLNRIHELLNGT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          100
                                                                                                                                                                                                                                                                                                                                                                                                     230
                                                                                                                                                                                                                                                                                                                                                                                                        240
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          120
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ----TTVHHVTL
                                                                                                                                                                                                                                                                                                                                                                                                                                                         180
                                                                                                                                                                                                                                                                                                                                                                                                     250
```

US-08-223-263-1 (1-353)

B.stearothermophilus DNA polymerase

16-UUN-1994 (first entry)
B.stearothermophilus DNA polymerase; thermal stability;
Bacillus stearothermophilus; DNA polymerase; thermal stability;
heat resistant; plasmid pUIF101. R45155 standard; Protein; 876 AA. R45155; 27-APR-1992; 131400. 27-APR-1992; JP-131400. (TAKI) TAKARA SHUZO CO LTD. Bacillus stearothermophilus (IAM11001). J05304964-A. .9-NOV-1993

WPI; 93-408323/51. N-PSDB; Q54170.

DNA polymerase gene - is isolated from plasmid pUIF101, useful as reagent for gene engineering research Example 2; Page 8-11; 11pp; Japanese.

DNA was extracted from Bacillus stearothermophilus IAM11001 and subjected to PCR amplification using the primers Q54168 and Q54169. A heat-resistant polymerase (R45155) can be recombinantly produced by transforming E.coli host cells with the gene comprising sequence by transforming $\bar{\text{E.coli}}$ host cells Q54170.

Initial Sequence 83 A; 55 48 I; 99 Score 876 AA; 25 N; 62 K; 49 22 K O 330 ኳጀ 2 37 P.C 38 37 11 11 ŝ 90 40 HE

50

Σ,

44 28

∀. છ 17 62

< H

Residue Identity 11 11 18 21**%** 69 Conservative Substitutions Optimized Score Matches 60 84 Mismatches Significance II 5.87

1

wit.

\$

Listing for Mary Hale

1

Tue Apr 11 13:27:36 1995

Ą

1

10.00

Page 12

US-08-223-263-1 (1-353) R28348 Bacillus caldotenax DNA polymerase

07-APR-1993 (first entry)
Bacilius caldotenax DNA polymerase I.
PolI; polymerase chain reaction; stra
plasmid pUI101. Misc_difference 1 R28348 standard; Protein; 877 R28348; Bacillus caldotenax. /note= "corresponds Location/Qualifiers to TTG codon" AA. strain YT-G;

DSM406;

27-MAY-1992; 304763.
03-JUN-1991; JP-157368.
07-NOV-1991; JP-318685.
24-FEB-1992; JP-072090.
25-FEB-1992; JP-073161.
06-APR-1992; JP-112400.
(TAKI) TAKARA SHUZO CO L EP-517418-A.

LTD.

```
Residue Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                DR N-PSDF, 23155.

DR N-PSDF, 23155.

DR N-PSDF, 23155.

PS Claim 2; Page 13-15; 30pp; English.

PS Claim 2; Page 13-15; 30pp; English.

PS Claim 2; Page 13-15; 30pp; English.

CG B. caldotenax YT-G was cultured and DNA was isolated. The DNA was complified by PCR using primers based on conserved amino acid motifs common the Poll sequences of E.coli, Phage T7, Thermus aquatious cand Streptococcus pneumoniae (see 31648 and Q31649). The 600bp amplified fragment was used as a probe to screen Southern blots camplified fragment DNA. Restriction mapping of three positive cof B. caldotenax genomic DNA. Restriction mapping of three positive cof Eragments (i.e. a 1.45kb HinclI fragment, a 2.4kb HindlII fragment cand a 2.1kb XhoI fragment) showed that the fragments overlapped. CC injusted with the vector pTV118N to produce plasmid pU1101 containing a ca. 3.5kb combined fragment. E.coli HB101 transformed containing a ca. 3.5kb combined fragment. E.coli HB101 transformed cult the recombinant plasmid pU1101 expressed a heat-resistant DNA polymerase deduced to have the amino acid sequence given here.
                                                                                                                                                                                                                                                                                                                                                                                                                       IGEKTAVKILROFGTVENVLASIHEIKGEKIKETLROHREMA---ILSKKLAAIRRDAPVELSLDDIAYQGE 200 210 220 230 240 250 260
                                                                                                                                                                             VVNEHGRFFLRPETALADPQFVAWLGDETKKKSMFDSKR--AAVALKW---
340 350 360 370 380
                                                                                                                                                                                                                                         V---PSRTSL-VLTLNELPNRTSGL-LET---NFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPG
                                                                                                                                                                                                                                                                                                      DREKVVALFKELGFQSF LEKMESPSSEEEKPLAKMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIA 270 280 290 300 310 320 330
                                                                                                                                                                                                                                                                                                                                            130 140 150 160 170
VRLLLGALQSILGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCV--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     HPLPTPVLL--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              GFEVKVISGDRDÍTGÍASÞHVTVDITKKGITDIEÞYTF-EAVREKYGITÞEQIVDIKGIMGDKSDNIÞGVÞG
130 X 140 150 160 170 180 190
FPL-PPTL--PTPVVQLHPL--LPDPSAPTPTPTSPLLNTSYTHSQNLSQEG
                                                                                                                    YLNRIHELLNGTRGLFPGPSRRTLGA-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            See also R28340-R28347 and Q31651.
Sequence 877 AA;
84 A; 54 R; 25 N; 48 D; 0 B; 2 C
48 I; 99 L; 62 K; 22 M; 33 F; 39 P
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Fujita K, Ishino Y, Kato I, Uemori
WPI; 92-408872/50.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Score
                                                          --LLLAAYLLDPAQGVDDVAAAAKMKQYEAVRPD-
400 410 420
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  X 10 20 30 40 MELTEL-LIVVMLLLT-ARLTLSSPAPPACDLRVLSKLLRDSHVLHSRL-
                                                                                                                                                                                                                                                              190
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       11 8
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   --PAVDFSLGEWK-TOMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQ
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Optimized Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              2 C;
39 P;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          80
                                                                                                                                                                                                                                                              210
                                                                                                                  -PDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                39
36
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          II
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                S O
                                                          -EAVYGKGAKRAVPDEPVLAEHLVRKAAAI
430 440 450
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        90
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  90
39
                                                                                                                                                                                                                                                              220
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        98
03
                    350
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ΗE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                o o'
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Mismatches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Significance
                                                                                                                                                                                                ---KGIEL-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        100
                                                                                                                                           290
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Σ.Ω
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                44
28
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Υ.Θ
                                                                                                                                                                                                  -CGVSFD----
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                18
62
                                                                                                                                                                               390
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      --SQCPEV--
                                                                                                                                                                                                                                                                                                                                                                -RRAP-PTTA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             11
                                                                                                                                           300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                < #
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             <u>ر</u>
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          .87
```

GDRERERGMEIS---

-PMCDKHTASVEKSQVGFIDYIVHPLWETWADLVHPYSAIRQSPSPPPEEESRGF

" Listing for Mary Hale

Tue Apr 11 13:27:36 1995

74

Initial Residue Complementary screening for genes and prods. - e.g. RAS protein and cAMP, that modify, complement or suppress genetic defect and carect associated phenotypic alteration correct associated phenotypic alteration.

To correct associated phenotypic alteration correct associated phenotypic alteration the SEQ ID NO. 2 and its control to SEQ ID NO. 3 1-55, i.e. the specification contains two sequences denoted as SEQ ID NO. 2 if the specification contains two sequences denoted as SEQ ID NO. 2 if the only reference to SEQ ID NO. 2 in the text is to a 10 base pair linker. The origin and identity of R14854 is therefore obscure. Other sequences in the specification were isolated from human glioblastoma cells and encode cyclic nucleotide PDEs and RAS-related polypeptides.

(They were isolated by their ability to complement or suppress genetic defects in a biochemical pathway involving cAMP or which is controlled by a RAS protein). 530 US-08-223-263-1 (1-353) R14854 Protein associated with biochemical 120 130 140 150 160 170 180 GQVRLLLGALQSLLGTQLPPQGRTTA--HKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPP--TTAVP ADLKTMVETKKV-TSSGVLLLDNYSDRIQ-LINTNSELALMYNDESVLENHHLAVGFKLL-470 480 490 HPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQ-Sequence : 80 A; 63 R; Wigler MH, Colicelli JJ; WPI; 91-339841/46. N-PSDB; Q14645. 20-APR-1990; WO9116457-A. Protein associated with biochemical pathway involving cAMP. RAS; oncogene; cancer; cyclic nucleotide phosphodiesterase R14854; R14854 standard; Protein; 900 30-JAN-1992 (first entry) (COLD-) COLD SPRING HARBOR. 19-APR-1991; U02714 Identity MELTELLLVVMLLLTARLTLSSPAPPACD-LRVLSKLLRDS--N 0 0 27 N; 36 D; 25 K; 26 M; 18 20% 74 Optimized Matches Conservative Substitutions 26 20 Ŧ, --QEDNCDIFQNLSKRQRQSLRKWVIDMVLATDMSKHMTLL 500 510 520 Ā 12 83 Score P C -VLRNMVHCA-DLSNPTKPLELYRQWTDRIMAEFFQQ 47 85 s o 68 45 100 82 HE pathway involv 40 Significance Mismatches 13 Ξ2 37 16 -HVLHSRLS-QCPEV ۲; G 27 49 11 11 < H S 5.87 239 0 18 H 590

Tue Apr 11 13:27:36 1995

Page 15

10.

US-08-223-263-1 (1-353) R12538 SMUC-87 intestinal mucin cDNA clone

02-SEP-1991 R12538 standard; Protein; 119 AA. SMUC; SMUC-87-intestinal mucin intestinal mucin; diagnosis; cancer. Afirst entry) cDNA clone

Homo sapiens. Region /label= repeat 51..73 28..50 Location/Qualifiers

/label= repeat W09108217-A. Region /label= repeat 98..119

/label= repeat

74..97

13-JUN-1991.
04-DEC-1990; U07087.
05-DEC-1989; US-447140.
(REGC) UNIV OF CALIFORNIA.
Kim YS, Gum JR;
WPI; 91-193147/26.

N-PSDB; Q12221.
Nucleic acid encoding human intestinal for producing prods. for diagnosis and mucin peptide(s) - used
treatment of disorders

such as cancer

homology segments of 24 amino acids and of 22 amino acids, demonstrating that these conserved segments may be of variable length, having amino acid insertions and deletions. The extra amino acid in the third repeat is a methionine (residue 85) while a threonine is absent from the fourth reconstraints. Disclosure; Fig 2(E); 57pp; English. Clone SMUC 87 contains 4 tandem repeats. homology segments of 24 amino acids and c

see also Q12217-21 and Q12247-51. Sequence 119 AA;

Ç G Ö -M 0 Z N ç; ш H;

X. Walter St. A. 30

15.6

Listing for Mary Hale

æ

Tue Apr 11 13:27:36 1995

1

Page 16

Residue 11. Initial SO 220 US-08-223-263-1 (1-353) R07674 Intestinal Ę TRITTINIVEREREPTGIQIPTERETITITIMVERERETITSQUERERETETE 60 70 80 90 100 100 LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHEL--Human intestinal mucin DNA, diagnosis and treatment of ε colitis. Epithelial cancer; cystic fibrosis; ulcerative glycosyl transferase; familial polyposis coli; glycosylation patterns. Disclosure; Fig 4E; 45pp; English.

The sequence was deduced from a clone isolated from a cDNA library using clone SMUC 41 as a probe. The DNA sequence allows determination of the primary amino acid structure, difficult to characterise biochemically because the protein backbone is so heavily glycosylated and conditions required to remove the carbohydrate tend to result in breakage of the protein. Prods. of the DNA can N-PSDB; Q06487 (REGC) University of California. Kim YS; Gum JR WPI; 90-348495/46. 25-JUL-1989; U03206. 14-APR-1989; US-338710 01-NOV-1990. Region /label= 22 Region R07674 standard; protein; 119 R07674; 290 WO9012892 Region Region Homo sapiens. /label= /label= /label= 22-FEB-1991 (first entry) Identity Score Į, 2 230 24 KFGL-23 23 aa ۲. aa repeat aa repeat 98..119 Intestinal mucin deduced ω repeat 74..97 repeat 51..73 <u>۲</u> 17 27% 21 28..50 Location/Qualifiers _ Optimized Matches Σ, Conservative QIIQKFKPIFFLSFTATTTTTENPTPTPTTTTTTTTTPTPTPTSTQSTTP G e.g. cancers, cystic fibrosis and Ŧ A A 23 320 Score ۳, Substitutions G from clone SMUC 87 --LNGTRGLFPGPSRRTLGAPDISSGTSDTGSLP s; 56 T; 500 30 35 colitis; Significance Mismatches 0 TTVTPTPTPTS-THT ξ 270 -TSPLLNTSYTHSQNLSQ 0 40 340 ۲; used Φ 11 11 H < 'n X 350 5 .45 73

X

1

69

Listing for Mary Hale

**

Tue Apr 11

13:27:36 1995

Page

Residue Initial 999999888 290 300 310 320 330 PNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVQLHPLLPDPSAPTPTP TPITTINTVTPTPTPTGTQTPTPTPTTTTTTMVTPTPTTTTSTQTPTPTPTTTTTTVTPTPTPTS-THT 60 70 80 90 100 110 X 20 230 240 250 260 270 280 LWQQGFRAKIPGLINQTSRSLDQIPGYLNRIHEL----LNGTRGLFPGPSRRTLGAPDISSGTSDTGSLP Sequence 1 A; 0 8 I; 2 be used to produce Abs userul for detecting the analysis of glycosylation patterns or glycosyltransferase activity. They can be used for the early detection and differential diagnosis of cancers esp. epithelial cancer, and also in diseases with altered intestinal mucin prodn. such as cystic fibrosis and colitis. See also R07670-76. Identity Score e 119 AA; R; 2 N; 0 L; 3 K; 1 KFGL---B B B 17 27% 21 QIIQKFKPIFFLSFTATTTTTENPTPTPTTTTTTTTTTTSTQSTTP 10 20 30 40 50 Matches Conservative K O Optimized 50 ΉŒ 23 Score P.C Substitutions ທ ປ 11 |1 S O 1 56 35 ΗE Significance Mismatches 00 Σ,2 TSPLINTSYTHSONLSO 0 N ۲; و 44 0 0 1 < # G X 350

ΞG

12. US-08-223-263-1 (1-353) R33420 Human IL-8

IL-8 receptor from clone p2

standard; Protein; 355 B

26-JUL-1993 (first entry)

Human II-8 receptor from clone p2. Interleukin-8 receptor; probes; ge intracellula-calcium mobilising; l gene therapy; gro receptor;
; ligand-binding; MIP-2 rece MIP-2 receptor.

Homo sapiens.

WO9306229-A.

01-APR-1993.

14-SEP-1992; U07641. 13-SEP-1991; US-759568

HEALTH HUMAN SERVICE

(USSH) US DEPT HE Murphy PM; WPI; 93-117549/14.

N-PSDB; Q38747

New interleukin-8 receptor aminoacid sequence - and corresp. cDNA expressed in Xenopus laevis oocytes or transfected host cells, for screening ligands of IL-8 receptor and gene therapy Claim 1; Fig 3; 39pp; English.

cDNA libraries from 2 and 3.5 kb fractions of poly(A)+ RNA from HL60 neutrophils sepd. of a sucrose gradient were made in UniZAP. The libraries were screened with F3R oligonucleotide probe (from rabbit IL-8 receptor) and under low stringency with a p2 cDNA probe synthesised from random primers, to isolated the clone p2, encoding human IL-8 receptor protein. The presence or absence of the DNA ecoding IL-8R or related MIP-2 receptor may be detected using portions of the p2 clone as probes. P2 may also be used to screen

37 4 3 1 4 Arti Rich 400 * 130-40

> also be used 18

for ligands of II-8R and may also patient deficient in II-8R. The intracellular calcium-mobilising Sequence 355 AA; 22 A; 18 R; 13 N; 13 D; 0 B; 9 25 I; 65 L; 13 K; 8 M; 21 F; 13 IL-8R is a gro rece and ligand-binding in gene therapy to a gro receptor and properties. treat has ø

ΑĊ 30 \$ 50 11 18 ı, E **6** О Σ. 17 14 Υ. ნ 25

< #:

Initial Residue Identity Score 11 II 17 23% 51 Conservative Substitutions Optimized Score 11 39 55 Significance Mismatches 11 0 0 S 125 125 0

QETCERRNHIDRALDATEILGILHSCINP 280 290 300 QLSGQVR---120 --LLLGALQSL---LGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPP 40 LIYAF IGQKFRH-310 150 -GLIKILAIHG-160 LISKDSLP

180 190 200 X 210 TTAVPSRTSLVLTINELPNRTSGLLETNFTASART -KDSRPSFV---GSSSGHTSTTL 340 350 X

13. US-08-223-263-1 (1-353) R28273 Sequence in D

low affinity recombinant human inter

standard; Protein; 360 æ

04-APR-1993 (first entry)

Sequence in a low affinity recombinant human interleukin-8

(IL-8) receptor polypeptide in 4AB

ното sapiens. W09218641-A. receptor polypeptide; G-protein-coupled

29-0CT-1992.

10-APR-1992; U02977. 10-APR-1991; US-685101. 09-JUL-1991; US-726606. 09-DEC-1991; US-803842.

(REPK) (UYBO-) UNIV BOSTON REPLICEN CORP.

KM, Witt

₽,

Navarro J, Thomas | WPI; 92-382123/46.

N-PSDB; Q30012.

Recombinant mammalian interleukin-8 receptor - used for screening interleukin-8 binding antagonists, used to treat inflammation Disclosure; Fig 2; 71pp; English.



Listing for Mary Hale

Tue Apr 11 13:27:36

1995

8888888888888888888888888 Sequence 360 . 22 A; 18 R; 14 : 25 I; 65 L; 13 : Rabbit high affinity II-8 receptor gene was isolated from rabbit peritoneal neutrophils and used as a source of poly(A)+ RNA, to produce a rabbit neutrophil cDNA library. 250,000 recombinant plaques were screened for those which hybridized to an antisense oligonucleotide (Q30015). This probe was designed based on the sequence derived from the second transmembrane domain of G-protein-coupled receptors. After tertiary screening, six plaques were isolated. The insert was sequenced. The protein deduced from the second transmembrane domain of C-protein-coupled receptors. After tertiary screening, six plaques were isolated. The insert was sequenced. The protein deduced from kb in size. This insert was sequenced. The protein deduced from the F3R clone demonstrates that it belongs to the family of G-protein-coupled receptors. The deduced protein sequence indicates seven putatuve transmembrane segments. A human peripheral blood leukocyte lambda gtl1 cDNA library (5' stretch) was screened with a 652 bp EcoRI/BamHI fragment (including nucleotides 27 to 625) of the rabbit F3R clone. After tertiary screening several human clones which hybridized to the rabbit IL-B probe were isolated. The insert of one such clone, termed 4AB was sequenced (030012). 360 AA; R; 14 N; 14 D; L; 13 K; 9 M; 22 ΉŒ 9 13 ьĊ 30 s o 12 18 ΗE 60 Σ, 17 14 **₹**€ 25 8 **<** #. was

Residue Identity Initial DRYLAIVHATRTLTQKRYLVKFICLSIWGLSLLLALPVLLFRRTVYSSNVSPACYEDMGNNTANWR--150 X 160 170 180 190 200 Score MELT-H B Conservative Substitutions Optimized Score 10 20 30 40
---ELLLVVMLLLTARLTLSSPAPPAC--DLRVLSKLLRDSHVLH B II 39 55 Significance Mismatches II łł 5.45 125 0

14.

US-08-223-263-1 (1-353)
R14838 Protein deduced from human Glioblastoma cell cDNA.

R14838 standard; Protein; 405 AA.

SWEDAC R14838;

30-JAN-1992 (first entry)

Protein deduced from human Glioblastoma cell cDNA

RAS; oncogene; cancer; cAMP. Homo sapiens.



Initial Score Residue Identity Complementary screening for genes and prods. - e.g. RAS protein and cAMP, that modify, complement or suppress genetic defect and cAMP, that modify, complement or suppress genetic defect and correct associated phenotypic alteration

Example 1B; Page 70; 169pp; English.

In the specification the sequence from which this amino acid sequence was deduced is given the SEQ ID NO. 13 and 15 described as the cDNA insert of plasmid pJC99. The insert sequence shows no significant homology to previously isolated genes. Plasmid pJC99 in E.coli (ATCC 68599) is described as the cDNA insert of plasmid pJC99 in E.coli (ATCC 68599) is described as the specification to particular sequences caution is advised in accepting that this sequence 405 AA; 30 A; 36 R; 2 N; 9 D; 0 B; 9 C; 25 Q; 4 I; 74 L; 10 K; 6 M; 7 F; 33 P; 40 S; (COLD-) COLD SPRING HARBOR. Wigler MH, Colicelli JJ; WPI; 91-339841/46. 31-OCT-1991. 19-APR-1991; U02714 20-APR-1990; US-511 containing a human glioblastoma cell cDNA insert encoding a RAS-related polypeptide. N.B. many of the references in the text of N-PSDB; Q14626. WO9116457-A. MELT-11 11 11 Optimized Score = 71
Matches = 90
Conservative Substitutions ELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLHSRLSQCPEV 25 E; 0 18 T; 2 sequence is indeed pJC99 ! are incorrect, therefore Significance = Mismatches = Σ. 32 G; 10 H; H 1) 5.45 196 0

LCVRRAPPTTAVPSRTSLVLTLNELP--

190

---RVAYQDPSSGCTSKT

330

HPLPTPVL----LPAVDFSLGEWKTQMEETKAQDILGAVTLL-

90

-LEGVMAARGQLGPTCLSSLLG

100

QLSGQVRLLLGA---LQSLLGTQ

120

130

```
X RGL
                   HTÖAALATLÄÄT
```

15.

Initial Score = Residue Identity = Gaps = = R N-PSDF Q26674.

R N-PSDF Q26674.

Prolypeptides similar to v-mlp protein of MPLV - for diagnosis and Prolypeptides similar to v-mlp protein of MPLV - for diagnosis and T treatment of myeloproliferative diseases

T treatment of myeloproliferative diseases

C claim 1; page 28; 75pp; French.

This polypeptide is encoded by clone MPLV107, isolated from a genomic bank prepared from clones of Mus dunni cells containing a unique copy of the MPLV provirus. The bank was screened with two 300bp long clones specific for MPLV env region. Restriction enzyme analysis showed that clone MPLV107 contained the entire MPLV genome except for the 3'LTR.

See R23971 and Q24674-7.

Sequence 635 Ah; 39 R; 12 N; 23 D; 0 B; 22 C; 41 Q; 38 E; 0 Z; 36 G; 19 H; 10 1; 78 L; 12 K; 8 M; 18 F; 60 P; 64 S; 39 T; 21 W; 19 Y; 28 V; US-08-223-263-1 (1-353) R23970 MPLV env p 50 60 70 80 90 100
LHSRLSQCPEVHP-----IP-----TPVILLPAVDFSLGEWKTQMEETKAQ-DILGAVTLLLEGVMAARGQLG KATVSDTCEEVEPSLLEILPKSSERTP--LP----LCSSQAQMDYRRLQPSCLGTMPLSVCPPMAESGSCC 560 570 580 590 600 610 19-OCT-1990; WO-F07623.

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

Charon M, Gisselbrecht S, Penciolelli JF, Souyri

Tambourin P, Varlet P, Vigon I, Wendling F;

WPI: 92-167154/20. 05-NOV-1992 (first entry)
MPIV env protein with growth factor
Myeloproliferative leukaemia virus;
Myeloproliferative leukaemia virus. Region 449..635 /note= "gp70 of MPLV" W09207074-A. R23970 standard; protein; 635 AA. R23970; 30-APR-1992. 9-OCT-1990; F00762 × MPIV env protein with growth factor receptor prope 17 28% 57 10 ocation/Qualifiers Optimized Score Matches Conservative Substitutions receptor properties. haematopoietic cell. 11 11 33 48 Mismatches Significance ĸ II 5 40 .45 62

20

()X

1. M. .

PT--CLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD

130

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

```
Query sequence being compared:US-08-223-263-1 (1-353)
Number of sequences searched: 75511
Number of scores above cutoff: . 4091
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    FastDB - Fast Pairwise Comparison of Sequences Release 5.4
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         > 0 < Of |O IntelliGenetics > 0 <
                                                                                                                                                                                                                                                                                                                                                                                                            F10000-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     U50000-+
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Results file sqlpir.res made by
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        100000-
                                                                                                                                                                                                                                   1000-
                                                                                                                                                                                                                                                                                                                                                          5000-
                                                     100-
                                                                                                                                                                                500-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : PIR 43, all entries
50-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            TTHIANHSYL-PLS-620 630
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            X
ADÖMÅ−∵
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 on Tue 11 Apr 95 9:51:38-PDT.
```

Tue Apr.11 13:27:36 1995

Page 23

PARAMETERS

Similarity matrix	Unitary	K-tuple	N
Mismatch penalty	Ľ	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	23		
Randomization group	0		
Initial scores to save	45	Alignments to save	15
Optimized scores to save	7e 0	Display context	10
	200	CENTOU CHAMICHTOC	

SEARCH STATISTICS

Number of residues: Number of sequences Number of scores ab	Times:	Scores:
of residues: of sequences searched: of scores above cutoff:	CPU 00:02:33.97	Mean 6
22468834 75511 4091		Median 7
	Total Elapsed 00:02:34.00	Standard Deviation 2.60

Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.
Cut-off raised to 12.

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

1. \$45331	Sequence Name
1. S45331 c-MPI ligand - human 353 353 353 133.36 0	Description
353	Init. Opt. Length Score Score Sig. Frame
353	Init. Opt. Score Scor
353	Opt. Score
353 353 133.36	sig.
0	Frame



Listing for Mary Hale

Tue Apr. 11, 13:27:36 1995

Page

1. US-08-223-263-1 (1-353) S45331 c-MPI ligand - human

ENTRY TITLE ORGANISM \$45331 #type complete c-MPI ligand - human #formal_name Homo sapiens #common_name man

Gaps Residue Identity Initial Score ACCESSIONS REFERENCE DATE SEQUENCE SUMMARY ABLITELLLIVVMLLLITARLITLSSPAPPACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEW
MELTELLLIVVMLLLTARLITLSSPAPPACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEW
MELTELLLIVVMLLLTARLITLSSPAPPACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEW
X 10 20 30 40 220 230 240 250 260 270 GLKWQQGFRAKIPGLLQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSSRTLGAPDISGTSDTGSLPPNL GLLKWQQGFRAKIPGLLUQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSRRTLGAPDISGTSDTGSLPPNL GLLKWQQGFRAKIPGLLQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSRRTLGAPDISGTSDTGSLPPNL 80 90 100 140

KTOMEETKAQDILGAVTLLLEGVMARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD

KTQMEETKAQDILGAVTLLLEGVMARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD

KTQMEETKAQDILGAVTLLLEGVMARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD #journal #title #accession #authors 220 ##residues ##status 230 11 11 10-Dec-1994 \$45331 \$45331 de Sauvage, F.J.; Hass, P.E.; Spencer, S.D.; Malloy, B.E.; Gurney, A.L.; Spencer, S.A.; Darbonne, W.C.; Henzel, W.J.; Wong, S.C.; Kuang, W.J.; Oles, K.J.; Hultgren, B.; Solberg Jr., L.A.; Goeddel, D.V.; Eaton, D.L. Nature (1994) 369:533-538
Stimulation of megakaryocytopoiesis and thrombopoiesis by the C-Mpl ligand. preliminary
1-353 ##label DEA
#length 353 #molecular-weight 37822 #checksum 1597 S4533 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change 90 353 100% 240 Optimized Score Matches Conservative Substitutions 100 250 110 260 353 353 120 Significance Mismatches 270 130 = 133.36 = 0 = 0 140

US-08-223-263-1 (1-353) S45330 thrombopoi thrombopoietin mouse

TITLE ORGANISM DATE ENTRY

\$45330 #type complete thrombopoietin - mouse #formal name Mus musculus #common name house 10-Dec-1994; #sequence_revision 10-Dec-1994; 10-Dec-1994 #text_change

ENTRY

A37232

mucin, tracheal (AMN-22) - human (fragment)

A37232 #type fragment
mucin, tracheal (AMN-22) - human (fragment)
#formal name Homo sapiens #common name man
14-Feb-T992 #sequence_revision 14-Feb-1992

#text_change

....

9

1 1 ×:

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

3. US-08-223-263-1	290 1.QPGYSPSPTHPP 1.QEGLPPSPSLAP 290 300	220 GLIKWQQGFRAKI + + GLISRLQGFRVK: 220 2:	150 PNAIFLSFQHLL PNALFLSLQQLLE 150	80 KTQMEETKAQDII 	X 10 WELTELLLVVMII WELTDLLLAAMLI X 10	Initial Score Residue Identity Gaps	*accession ##status ##residues SUMMARY SEQUENCE		ACCESSIONS REFERENCE #authors
-1 (1-353)	290 300 310 320 330 340 350 X LQPGYSPSPTHPPTGYTLFPTLFTPVVQLHPLLDDPSAPTPTPTSPLLNTSYTHSQNLSQEG	220 230 240 250 260 270 280 GLIKWOQGFRAKI-PGLINQTSRSLDQLPGYLNRIHELINGTRGLFPGPSRRTIGAPDISSGTSDTGSLPPN	150 200 210 PNAIFLSFQHLLRGKVRFLMIVGSSTLCVRRAPPTRAVPSRTSLVITLNELPNNTSGLIETNFTASARTTGS	80 90 100 120 130 140 KTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD	X 10 20 30 40 50 60 70 MELTELLIVVMLILITARLIISSPAPPACDLRVISKLIRDSHVIHSRLSQCPEVHPLPTVVILPAVDFSIGEW	= 194 Optimized Score = 252 Significance = 72.25 = 71% Matches = 255 Mismatches = 97 = 5 Conservative Substitutions = 0	54330 preliminary s 1-356 ##label LOK #length 356 #molecular-weight 37835 #checksum 9983	O'Hara, P.J.; Hagen, F.S.; Roth, G.J.; Foster, D.C. O'Hara, P.J.; Hagen, F.S.; Roth, G.J.; Foster, D.C. Nature (1994) 369:565-568 Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo.	S45330 S45330 Lok, S.; Kaushansky, K.; Holly, R.D.; Kuijper, J.L.; Lok, S.; Kaushansky, K.; Holly, R.D.; Kuijper, J.L.; Lofton-Day, C.E.; Oort, P.J.; Grant, F.J.; Heipel, M.D.; Burkhead, S.K.; Kramer, J.M.; Bell, L.A.; Sprecher, C.A.; Blumberg, H.; Johnson, R.; Prunkard, D.; Ching, A.F.T.; Mathewes, S.L.; Bailey, M.C.; Forstrom, J.W.; Buddle, M.M.;

200 Com 200 Co

> Page 27

Residue Identity Initial Score ORGANISM DATE ENTRY ACCESSIONS REFERENCE SEQUENCE **DOMMARY** TITLE US-08-223-263-1 (1-353 A29746 #authors
#journal
#title SLLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAP---60 x 70 80 90 100 110 120 LPTPVLLPAVDFSLGEMKTQ-MEETKAQDILGAVTLLLEGVMAARGQLGFTCLSSLLGQLSGQVRLLLGALQ #authors Gerard, C.; Eddy Jr., R.L.; Shows, T.B.
#journal J. Clin. Invest. (1990) 86:1921-8
#title The core polypeptide of cystic fibrosis tracheal mucin
contains a tandem repeat structure. Evidence for a common
mucin in airway and gastrointestinal tissue.
#cross-references MUID:91072667
#accession A37232 TSTQTPTPTPITT LINGTRGLFPGPSRRTLGAPDISSGTSDTGSLPFNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHP--200 210 220 230 240 LIBELPIRTSGLLET-NETASARTTGSGLLKWQQGFRAKIPGLLNQT----SRSLDQIP-------LLPDP-##molecule_type mRNA ##residues 1-29 ##cross-references GB:M57417 Y_ #length 294 #checksum 1596 ##status 260 190 TPTPTFTGTQTPTPTPTTTTTTMVT-X 10 20 epidermal 0 0 0 30-Sep-1993 A37232 A37232 TVTPTPTPTSTØRTTPTSITTTTVTPTPT #formal name Mus musculus #common_name house mouse
08-Mar-1989 #sequence_revision 30-Jun-1989 #text_change A29746 #type complete epidermal growth factor-binding protein type A precursor mouse APTETETS-60 26 21**%** 61 -PISI-----TTTTTVTPTPTPTTTTTTTTTTTTTTTTTTSTSTTTTV-T 60 70 80 90 100 1-294 ##label GER preliminary growth factor-binding Optimized Score Matches taa Conservative Substitutions 340 350 X 160 290 220 170 ----PTPTITSTQTPTPTPITTTTVTPTP protein type A pre 300 55 70 Significance Mismatches 30 102(6) 180 PTTAVPSRTSLVLT 11 II II --GYLNRIHE 7.69 190 0

1 ** ig. N.

ENTRY

US-08-223-263-1 (1-353)

PQ0452

extensin-like protein - Persian tobacco (strain

ruu432 #type fragment extensin-like protein - Persian tobacco (strain S2S3) (fragment)

#formal name Nicotiana alata #common name Persian tobacco 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993

ORGANISM

ACCESSIONS

4

10,10

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

```
Gaps
                                                                                                                                                                                                                                                                                                                                                                                                                       Initial Score
Residue Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    FEATURE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ACCESSIONS
REFERENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     CLASSIFICATION
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   SUMMARY
                                                                                    NPNDLQCVSIKLHPNEVCVKAHILKVTDVMLCAGEMNGGKDTCKGDSGGPLIC---170 180 190 200 210
                                                                                                                                                                                                                                                                                                       KNSQPWQVAVYYLDEYLCGGVLLDRNWVLTA--
40 X 50 60
                               200 210 220 230 240 LNELPNRTSGLIETNFTASA-RTTGSGLIKWQQGFRAKIPGLINQTSRSL
                                                                                                                                        130 140 150 160 170 180 190 LIGALQSILIGTQLPPQGRTTAH-KDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLT
                                                                                                                                                                                               HPDFNMSLLOSVPTGADLSNDLMLLRLSKPADLTDVV--KPIDLPTTEPKLGSTCLASGWGSINQLIYQ---
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         #authors Drinkwater, C.C.; Evans, B.A.; Richards, R.I.
#journal Biochemistry (1987) 26:6750-6756
#title Mouse glandular kallikrin genes: identification
characterization of the genes encoding the epid
factor binding proteins.

#cross-references MUID:88107594
                                                                                                                                                                                                                                               HPLPTPVLLPAV--DFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLG----QLSGQVRL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   25-251
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             #accession
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              #contents
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            ##residues
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ##molecule_type DNA
                                                                                                                                                                                                                                                                     8
                                                                                                                                                                                                                                                                                                                                          X 10 20 SV MELTELLLVVMIL-----LTARLTLSSPAPPACD--LRVLSK--LLRDSHVLHSRLSQCPEV
                                                                                                                                                                                                                                                                                                                                                                                                              11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     #superfamily trypsin; trypsin homology
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             A29746
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 #domain trypsin homology #label TRP
#length 259 #molecular-weight 28384 #checksum
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Strain BALB/c, salivary gland
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      A90522
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         28-Oct-1994
-APAIYTKLIKFTSWIKDTMAKNP
                                                                                                                                                                                                                                                                                                                                                                                                              23
23
45
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      1-259 ##label DRI
the authors translated the codon
                                                                                                                                                                                                                                                                                                                                                                                                            Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                           Optimized Score
Matches
                                                                                                                                                                                                                                                                     80
                                                                                                                                                                                                                                                                  90
                                                                                                                                                                                                                                                                                                                     -AHCYEDKYNIWLGKNKLFQDEPSAQHRLVSKSFF
                                                                                                                                                                                                                                                                                                                                                                                                                                n n
                                                                                                                                                                                                                                                                    100
                                                                                                                                                                                                                                                                                                                                                                                                                             59
59
                                                                                                                                                                                                                                                                                                                                                                                                                           Significance = Mismatches =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        for
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              tion and epidermal
                                                                                      -DGVLQGITSWGST
220 230
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   2468
                                                                                                                                                                                                                                                                                                     90
                                                                                                                                                                                                                                                                                                                                                                                                          6.53
142
0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        210
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                growth
                                                                                                                                                                                                                                                                120
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          as
```

Tue Apr 11 13:27:36 1995

Page 29

##residues
SUMMARY
SEQUENCE Initial Score Residue Ident: Residue Identity Initial Score ACCESSIONS REFERENCE ENTRY TITLE ORGANISM SUMMARY SEQUENCE REFERENCE US-08-223-263-1 (1-353) S45222 ferrichrome-iron transport protein -RDAHGWTWASGALLEDLMPWRWPRIMAALFAGVML---AVAGCIIQRLTGNPMASPEVLGISSGAAFGVVLM 370 380 390 400 410 420 430 300 310 320 330 340 350 X
GYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTTSPLLNTSYTHSONLSQEG #journal #title PPSPSPPPPSPSPPPPSPSPP-PPSPPADDMAPSPSP-AAAPTPPDMLLRPSARLAGNMVISGLH 60 LPTPVLLPAVDF #accession #submission #authors FRAKIPGLINGTSRSLDQIPGYLNRIHELLNGTRGLFP--#accession *contents fauthors ##status preliminary ##residues 1-59 ##label FUJ ##cross-references EMBLID26562 ##cross-references EMBLID26562 Y #length 659 #molecular-weight 70335 Identity MELTELL-----LVVMLLLTARLTLSSPAPPACDLRVLSKLLRDS-HVL-HSRLSQCPEVHP HGTESL--PDATLQR-FKFKWPFFGKSPKNSPKSSPSRSPPPKREQPSPPPPVKSPPPPSPP X 10 20 30 40 50 H O II II II Specific expression of an extensin-like gene in the style of Nicotiana alata.
Style, Strain S2S3
PQ0452 ferrichrome-iron transport protein - Escherichia coli
#formal name Escherichia coli
07-0ct-1994; #sequence_revision 07-0ct-1994; #text_change submitted to the EMBL Data Library, January 1994 \$45222 S45222 Fujita, N. \$45181 07-0ct-1994 22 24% 77 22 28% 11 80 70 110 120 -SIGEWKTQMEETKAQDIIGAVTLILIEGVMAARGQIGPTCISSIIGQISGQVRI Matches = 35 Conservative Substitutions Optimized Score Matches = 79 Conservative Substitutions Optimized Score #type complete 260 208 11 11 100 ---GPSRRTLGAPDISSGTSDTGSLPPNLQP 59 79 30 35 Escherichia c Mismatches Mismatches Significance = Significance = #checksum 1419 6.15 170 0 6.15 79 0

Listing for Mary Hale

Tue Apr:11 13:27:36 1995

Page 30

```
7. US-08-223-263-1 (1-353)
S07318 ferrichrom
                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ACCESSIONS
REFERENCE
#authors
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ORGANISM
DATE
 Initial Score
                                                                                             KEYWORDS
FEATURE
1-20
21-659
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    ENTRY
TITLE
                                                                                                                                                                     #map_position
#start_codon
CLASSIFICATION
                                                                                                                                                                                                                                                    GENETICS
                                       SEQUENCE
                                                           SUMMARY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       MALTPTRIAL---
560
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           180 190 200 210 220 230 240 250 TTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRI
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      SALVGGLLLVFADWCGRMVLFPFQIPAGLLSTFIGAPYFIYLLRKQSR
620 630 640 650 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         HELLNGTRGLF-PGPSRRTLGAPDISSGTSDTGSLPP---NLQPGYSPSPTHPPTGQY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      LLMMLQ-ASGDPRMAQVLTWISGSTYNATDAQVWRTGIVMVILLAITPLCRRWLTILLPL-GGDT---ARAVG 500 510 520 530 540 550
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              LLGALQSILGTQLPPQ-----
                                                                                                                                                                                                                              #gene
                                                                                                                                                                                                                                                                                                                                                                    #accession
                                                                                                                                                                                                                                                                                                                                                                                                                                               #title
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          L---FIVPGNAFGWLLPAGSIG
                                                                                                                                                                                                                                                                                                                                                                cross-references MUID:87014116
                                                                                                                                                                                                                                                                                                                                                                                                                                                               # journal
                                                                                                                                                                                                                                                                                                                              ##molecule_type DNA
##residues 1-65
                                                                                                                                                                                                                                                                                                        ##cross-references EMBL:X04319
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ferrichrome-iron transport protein fhuB precursor
   ti
                                                                                                                                                                                                            fhuB
4 min
                                                                                                                                                                                                                                                                                                                                                                                                 Koester, W.; Braun, V.
Mol. Gen. Genet. (1986) 204:435-442
Iron hydroxamate transport of Escherichia coli: nu sequence of the fhuB gene and identification of protein.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ferrichrome-iron transport protein fhuB precursor -
Escherichia coli
#formal name Escherichia coli
31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
18-Jun-1993
                                                                                                                                                                                           GIG
                                                       #length
                                                                                                                                                                   #superfamily vitamin B12 transport protein btuC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       S07318
                                                                                                                                                   transport; membrane protein
                                                     #domain signal sequence #status predicted #label SIG\
#product ferrichrome-iron transport protein fhuB #status
predicted #label MAT

pth 659 #molecular-weight 70335 #checksum 1419
                                                                                                                                                                                                                                                                      the authors translated the codon ATC for residue 311 as {\tt Thr}
                                                                                                                                                                                                                                                                                                                            1-659 ##label KOE
 22
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ---LLIAACLTATATMTIGPL--SFVGIMAPHIARMMGFRRTMPHI-
570 580 590 600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                -GRTTAHKD--
Optimized Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    #type complete
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      150 160 170
--PNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPP
 ()
 59
Significance =
                                                                                                                                                                                                                                                                                                                                                                                                                                             nucleotide
 6.15
```

Tue Apr 11:13:27:36 1995

Page 31

ENTRY TITLE QYBYP pyruvate carboxylase (EC 6.4.1.1) 1 - yeast (Saccharomyces cerevisiae) ALTERNATE NAMES pyruvate carboxylase ORGANISM ACCESSIONS AL2633; SO5760; A29722 REFERENCE #authors #itile #cross-references mand domain structure of yeast pyruvate carboxylase. ##cross-references EMBL:J03889 ##cross-references EMBL:J03889 REFERENCE #authors ##cross-references EMBL:J03889 REFERENCE #oria, C.P.; Lim, F.; Wallace, J.C. ##across-references EMBL:J03889 REFERENCE #oria, C.P.; Lim, F.; Wallace, J.C. #authors #across-references EMBL:J03889 REFERENCE Morris, C.P.; Lim, F.; Wallace, J.C. #authors #across-references EMBL:J03889 REFERENCE Morris, C.P.; Lim, F.; Wallace, J.C. #authors #across-references EMBL:J03889 REFERENCE S05760 #accession S05760 #accession S05760	260 270 280 290 X 300 HELLNGTRGLF-PGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQY	130 140 150 160 170 140 11	Column	tches = 79 Mismatches = nservative Substitutions =
---	--	----------------------------	--------	--

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

Page

```
Initial Score = Residue Identity = Gaps =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         SUMMARY
SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            FEATURE
157-331
353-468
569-908
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Initial Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                #map_position
CLASSIFICATION
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            GENETICS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         KEYWORDS
                                                                                                                                                                                                                                                                                           PNIPFQMILRGANGVAYSSIPDNAIDHFVKQAKDNSVDIFRVFDAINDLEQIKVGVDAVKKAGGVVEATVCF
630 640 650 660 670 680 690 700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ##residues 1003-1178 ##label MOR ##cross-references EMBL:J03889 #accession A29722
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              KLKSNPSVPHLHDAQGNVINVTKSAPPSGWRQVLLEKGPAEFAR-----
510 520 530 540
                                                                                                                                                                                                                                                                                                                                                                                                                                           DTTWRDAHQSLLATRVRTHDLATIAPTTAHALAGAFALECWGGATFDVAMRFLHEDPWQRLRKLR----SLV 570 580 590 600 610 620
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   RIRGVKTNIPFLLTILTNPVFIEGTYWGT-FIDDTPQLFQMVSSQNRAQKLLHYLADVADNGSSIKGQIGLP
440 X 450 460 470 480 490 500
-HDSAGTRVASMTACALAGADVVDVAINSMSGL--TSQPSINALLASLEGNIDTGINVEH
770 780 790 800 810 X 820
                                                            310 320 330 340 350 X
LFPLPPTLPTPVVQLHPLLPPSAPTPTPTSPLLNTSYTHSQNL--SQEG
                                                                                                                                             SGDMLQPGKKYNLDYYLEIAEKIVQMG--THILGIKD-MAGTMKPAAAKLLIGSLRAKYPDLPIHVHT----
710 720 730 740 750 760
                                                                                                                                                                                                          SLD-QIPG-YLNRIHELLNGTRGLFPGPSRRTLGAPDISSGT---SDTGSLPPNLQPGYSPSPTHPPTGQYT
                                                                                                                                                                                                                                                                                                                                                            60 70 110 120
FOULDAYDF SLGEWKTOMEETKAQDILGAVTILLEGVMA--ARGQLGPTCLSSLLGQLSGQVR-----LL-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   1089-1178
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   gene
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ---LGALQSILGT-----QIPPQGRTTAHKDPNAIFISFQHLLRGKV--RFLMLVGGSTLCVRRAPPTTAV
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   ##molecule type protein
##residues 1124-1149 ##label MOR2
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ##molecule_type DNA
##residues_____100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        MELTELLLVVMLLLTARLTLSSPAPPACDLRVLS----KLLRDSHVLHSRLSQCPEVHPLP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                #domain ATP/bicarbonate-binding #label ATB1\
#domain ATP/bicarbonate-binding #label ATB2\
#domain pyruvate-binding #label PYR\
#domain biotinyl- or lipoyl-binding #label BIO\
#binding site biotin (Lys) (covalent) #status predicted
#length 1178 #molecular-weight 130098 #checksum 6631
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               LISTA:PYC1
7
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    #superfamily pyruvate carboxylase
biotin; gluconeogenesis; homotetramer; ligase; zinc
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Optimized Score = Matches =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        68 Significance = 92 Mismatches =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      --QVRQFNGTLLM
560
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      u u
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 6.15
229
0
```

Initial Score Residue Identity #map position a CLASSIFICATION # KEYWORDS FEATURE 9. US-08-223-263-1 (1-353) 1136 SUMMARY SEQUENCE ENTRY TITLE REFERENCE #authors GENETICS REFERENCE ACCESSIONS ORGANISM ALTERNATE_NAMES RIRGVKTNIPFLLTLLTNPVFIEGT-YWTTFIDDTPQLFQMVSSQNRAQKLLHYLADLAVNGSSIKGQIGLP #authors Stucka, R.; Dequin, S.; Salmon, J.M.; Gancedo, C.
#journal Mol. Gen. Genet. (1991) 229:307-315

#title DNA sequences in chromosomes II and VII code for pyruvate carboxylase isoenzymes in Saccharomyces corevisiae:
analysis of pyruvate carboxylase-deficient strains.
#cross-references MUID:92017667 #accession #submission TPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLL-----LGA *accession 6 gene ##molecule_type DNA ##residues 1-1180 ##label DUB ##cross-references EMBL:Z36087 ##cross-references EMBL:X59890 X 10 20 30 40 50 MELTELLLVVMLLITARLITLSSPAPPACDLRVLS-----KLLRDSHVLHSRLSQCPEVHPLP U 11 H pyruvate carboxylase (EC 6.4.1.1) 2 - yeast protein YBR1507; protein YBR218c
#formal name Saccharomyces cerevisiae
26-Aug-T994 #sequence_revision 09-Sep-1994 #text_change
14-Sep-1994
S46094; S17469
\$45782 Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, Pierard, A.; Scherens, B.; Vierendeels, F. submitted to the Protein Sequence Database, August 1994 #superfamily pyruvate carboxylase
biotin; gluconeogenesis; homotetramer; ligase; zinc S46094 #type complete pyruvate carboxylase (EC 6.4.1.1) 2 -S46094 S17469 #length 1180 LISTA: PYC2 23 23 65 80 Optimized Score Matches Conservative Substitutions 11 11 ----KGPSEFAKQVRQFNG--TLIMDTTWRDA 550 560 91 91 Significance Mismatches 110 yeast (Saccharomyces (Sacch 11 6.15 237 0 Ή :

Listing for Mary Hale

Page

Residue Identity REFERENCE #authors #journal [nitial DATE ENTRY SUMMARY ACCESSIONS ORGANISM #authors Jackson-Grusby, L.; Kuo, A.; Leder, P.
#journal Genes Dev. (1992) 6:29-37
#title A variant limb deformity transcript expressed in the
embryonic mouse limb defines a novel formin.
#cross-references MUID:92112033 #accession ##status ##cross-references EMBL:X62379
#length 1206 #molecular-weight 133463 #checksum ##residues Score 11 II II formin isoform IV - mouse formin isoform IV - mouse
#formal name Mus musculus #common name house mouse
19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change S24407 19-Feb-1994 preliminary 1-1206 ##label JAC Optimized Score Matches Conservative Substitutions #type complete 67 95 Significance = Mismatches = 501 6.15 220 0

MELTE----

VHPLPTPVILLPAVDFS-LGEWKTQMEETKAQDILGAVTILLLEGVMAARGQLGPTCLSSILGQLSGQVRLLLG

100

ALQ----SLLGTQLPPQGRTTAHKDPNAIFLSFQHLLR-GKVRFLMLVGGSTLCVRRAPPTTA-VPSRTS

150

160

PKKGSTADTSELEALKRKMKHEK--ESLRAV-

--FERSKSRPADSPSDPKSPDQSPTEQDDRTPG

450

Listing for Mary Hale

Tue Apr. 11 13:27:37 1996

Residue Identity Initial Score ACCESSIONS REFERENCE #authors ENTRY TITLE ORGANISM 11. US-08-223-263-1 (1-353) SADTLEPSSTTKVTETKGASPTSLRÅSQTWLVSEEASEKGLGPEKITAPPQHQLPPGIASEGFPC-DNFKEQ 570 580 590 600 610 620 630 190 LVLTLNE----#journal #title #accession S11515 FPLPPTLPTPVV--QLHPLLPDPSAPTPTPTS-250 260 270 280 290 290 300 CYLNRIHELLNGTRGLFPGPSRRTLGAP-DISSGTS----DTGSLF-----PNLQPGYSPSPTHPPTGQYTL RLQAVWPPPKTKDTEEKVGLKYTEAEYQAAI----LHLKREHKEEIETLQAQFELKTFHIRGEHALVTAR--60 ##status preliminary ##residues 1-1468 ##label WOY ##cross-reference EMBL:X53599 ##cross-reference EMBL:X53599 ---VPKKLTISLTQLSPSKDSKDIHAPFQTREGTSSSSQQKISPPAPPTPPPLLPPPLIPPPPLPPG----L LEEAIENIKQQIEKRREGCEEMRDVCISTDDDDCSPKAFRNVCIQTDRETFIKPCDAESKATRS-SQI 540 550 560 570 580 590 600 X 10 20 30 40 MELTELLLVVMLLLTARLT-LSSPAPPACDLRVLSKLLRDSHVL-70 ---LPNRTSGLLE-TNFTASARTTGS---GLLKWQQGFR---AKIPGLLNQTSRSLDQIP formin - mouse S11515 Woychik, R.P.; Maas, R.L.; Zeller, R.; Vogt, T.F.; Leder, P. Nature (1990) 346:850-853
'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. #formal name Mus musculus #common name house mouse 22-Jan-1994; #sequence_revision 22-Jan-1994; #text_change S11515 formin - mouse 22-Jan-1994 490 22 23* 86 80 Conservative Substitutions Matches Optimized Score #type complete 630 90 (1 II 510 100 340 350 X --PLLNTSYTHSQNLSQEG 67 95 Mismatches Significance 110 HSRLSQCPEVHPLP 120 H II 6099 6.15 224 0

12. US-08-223-263-1 (1-353) S28148 erythropoi erythropoietin -

ENTRY TITLE ORGANISM DATE REFERENCE ACCESSIONS #authors Nagao, M.; Suga, H.; Okano, M.; Masuda, S.; Ikura, K.; Sasaki, R. Biochim. Biophys. Acta (1992) 1171:99-102 \$28148 #type complete erythropoietin - rat #formal name Rattus norvegicus #common name Norway rat 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change S28148 22-Nov-1993 Narita,

Η.;

#accession ##status title ##residues #length 192 #molecular-weight 21286 #checksum 4587 Nucleotide sequence of rat erythropoietin S28148 preliminary 1-192 ##label NAG

#journal

Residue Identity Initial Score

H H

21 27**%** 24

Optimized Score Matches

H 9

53

Significance = Mismatches =

11 11

5.77 118 0

Conservative Substitutions

SEQUENCE

Tue Apr 11 13:27:37 1998

1

Page

జ్ఞ

250	170 180 190 200 210 220 230 240 TLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIFGLLNQTSRSLD	Initial Score = Residue Identity = Gaps = The score =	SUMMARY SEQUENCE	2-131 132-215 216-468	#gene CLASSIFICATION KEYWORDS FEATURE	##molecule_type_DN ##residues 1 ##cross-references		rs	ORGANISM DATE ACCESSIONS	ALTERNATE_NAMES	ENTRY	13. US-08-223-263-1 FOMVMU ga	130 140 150 160 170 180 LQS-LLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPPTTAVPSRTS	70 80 100 110 120 SLGEWKTQMEETKAQDIIGAVTLLLEGVMAARGGJGPTCLSSLLGQLSGQVRLLLGA NFYANKFMEVEEQAVEVWQGLSLLSBAILQAQALQANSSQPPESIQHHIDKAISGLRSLTSLLR-VIGAQKE 80 90 100 110 120 130
260	0 190 TAVPSRTSLVLTL	21 22 % 15	P30 length 468	#pro P1	gag #superfam coat prot	type DNA 1-46 erences G	Moloney murine nucleotide s	A42745 Huai, L.; Bishop,	#formal name 30-Jun-1993 # 08-Apr-1994 A42745	core poly	FOMVMU #type gag polyprotein MuSVts110)	}-1 (1-353) gag polyprotein	140 RTTAHKDPN RTLTADTFC 160	90 .QDILGAVTL VEVWQGLSL 90
270	200 NELPNRTSG	Optimized Sco Matches Conservative		<pre>#product core #product inne P12\ #product core</pre>	ily mamma ein; core	DNA 1-468 ##label es GB:M96854			name Molon 993 #seque -1994	0, 2		t	150 AIFLSFQHL: KLFRVYSNF: 170	LLEGVMAA
280	00 210 220 230 TSGLLETNFTASARTTGSGLLKWQQGFRAKIP -DHWKDVERIAHNQSVDVKKRRWVTFCSAEWP 20 30 40	Score = 35 = 40 ive Substitutions	5	<pre>core protein pl5 inner coat prote core shell prote</pre>	<pre>gag #superfamily mammalian retrovirus gag polyprotein coat protein; core protein; polyprotein</pre>	ноа	virus	Chiocca, S.M.; Gilb	#formal name Moloney murine san 30-Jun-I993 #sequence_revision 08-Apr-1994 A42745	core shell	complete - Moloney muz	Moloney murine	160 LRGKVRFLMLV LRGKLKLV 180	ALQANSSQPPI
290	220 RTTGSGLLKW QSVDVKKRRW 30		52681	#statu in p12 in p30	virus gag polyprotei		MuSVts110 ne expres	Gilbreth, M.A.	sarcoma virus ion 30-Jun-1993	protein p	murine sarcoma	ine sarcoma	170 FIMIVGGSTLCVRRAP 	100 -RGQLGPTCL ESLQLHIDKA 120
300	230 QQGFRAKIPG VTFCSAEWP- 40	Significance Mismatches	#checksum	a	polyprotei n		DNA: cloning sion.	.; Ainsworth,	rus 1993 #text	p30; inner	virus	a virus (st	180 APPTTAVPSR DR X	100 120 -RCQLGPTCLSSLLCQLSGQVRLLLGA -RCQLGPTCLSSLLCQLSGQVRLLLGA
310	240 LLNQTSRSLD TFNVGW 50	= 5.77 = 120 = 0	3522	<pre>predicted #label predicted #label</pre>	H		ing,	th, J.R.;	_change	coat protein	(strain	r	IS	120 VRLLIGA LR-VLGAQKE 140

/<u>.</u> .

14. US-08-223-263-1 (1-353) S22701 mannosyltr Residue Identity #gene KEYWORDS FEATURE 16-30 Initial Score SUMMARY SEQUENCE ACCESSIONS REFERENCE ORGANISM DATE ENTRY TITLE GENETICS 320 330 340 350 X
PT-PVVQLHPLLEDPSAPTETETESLINTSYTHSQNLSQEG PRDGTFNRDLITQVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPFWVKPFVHPKPPPPLLPSAPSLPLEPPL 60 70 80 90 100 110 120 203,281,341,393 STPPQSSLYPAL-TPS-LGAKPKPQVLSDSGGPLIDLLTEDPPPYRDPRPP 130 140 150 160 170 QIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTL 20 30 70 80 80 TISSPAPPACDLRVLSKL--LRDS--HVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEET--KAQDI #journal EMBO J. (1992) 11:2511-2519
#title OCH1 encodes a novel membrane bound mannosyltransferase:
 outer chain elongation of asparagine-linked
 oligosaccharides.
#cross-references MUID:92331603
#accession S22701 ---NLKKQITVNKKKNQ-----LHNLRDQLSFAFPYDSQAPIPQRVWQTWKVG--ADDKNFPSSFRTYQKTW
70 80 90 100 110 120 #authors ##molecule type DNA ##residues 1-480 ##label NAK ##cross-references EMBL:D11095 MSRKLSHLIATRKSKTIVVTVLLIYSLLTFHLSNKRLLSQFYPSKDDFKQTLLPTTSHSQDIX 10 20 30 40 50 60 11 (1 1) mannosyltransferase (EC 2.4.1.-) - yeast (Saccharo #domain transmembrane #status predicted #label TMM\
3 #binding site carbohydrate (Asn) (covalent) #status
predicted
predicted
#length 480 #molecular-weight 55155 #checksum 9502 LISTA:OCH1 glycoprotein; transmembrane protein #formal name Saccharomyces cerevisiae
04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
09-Sep-1994 S22701 #type complete mannosyltransferase (EC 2.4.1.-) -Nakayama, K.I.; Nagasu, T.; Shimma, Y.; Kuromitsu, J.; Jigami, Y. cerevisiae) 21 20% 43 Optimized Score = 56 Matches = 73 Conservative Substitutions yeast 5.77 236 0

...

15. US-08-223-263-1 (1-353) FOMVM gag polypr

gag polyprotein - Moloney murine sarcoma virus

ACCESSIONS REFERENCE ENTRY TITLE ORGANISM CONTAINS #authors Reddy, E.P.; Smith, M.J.; Aaronson, S.A.
#journal Science (1981) 214:445-450
#title Complete nucleotide sequence and organization of the Moloney
murine sarcoma virus genome.
#cross-references MUID:82039559 FOMVM #type complete
gag polyprotein - Moloney murine sarcoma virus
core protein p15; core shell protein p30; inner coat proud p12; nucleoprotein p10
#formal name Moloney murine sarcoma virus
18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change
30-Sep-1933 A94261; B00645; A03932 protein

contents Provirus

#accession #molecule_type genomic RNA 1-538 ##label RED A94261

REFERENCE fauthors ##residues A00645

#title Nucleotide sequence of the genome of a murine sarcoma virus. #cross-references MUID:82115347 #contents Clone 124, circular #journal Cell (1981) 27:97-108 Van Beveren, C.; van Straaten, F.; Galleshaw, J.A.; Verma, I.M.

##molecule_type DNA B00645

GENETICS ##residues 1-518, 'K', 520-538 ##label VAN

#gene CLASSIFICATION FEATURE #superfamily mammalian retrovirus gag polyprotein I

Listing for Mary Hale

Tue Apr-11 13:27:37 1995

8

Gaps Residue Identity FastDB - Fast Pairwise Comparison of Sequences Release 5.4 Initial Score SUMMARY SEQUENCE 2-131 132-215 216-478 479-534 VGWPRDGTFNRDLITQVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPPWVKPFVHPKPPPPLLPSAPSLPLE
50 60 70 80 90 100 110 0 180 230 230 220 220 220 TICVERAPPTTAVPSRTSLVLTIMELPNRTSGLLETNFTASARTTGSGLLK---WQQGFRAKIPGLLNQTSR 350 X PTLPT-PVVQLHPLLPDPSAPTPTPTSPLLNTSYTHSQNLSQEG SLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLP IntelliGenetics MGQTVTTPLSLTL--DHWKDVERLAHN--X #product core protein p15 #label P15\
#product inner coat protein p12 #label P12\
#product core shell protein p30 #label P30\
#product nucleoprotein p10 #label P10
#length 538 #molecular-weight 61209 #checkanm ' 21 24% 21 260 Optimized Score = 36 Matches = 43 Conservative Substitutions --QSVDVKKRRWVTFCSAEWP----TFN 30 290 Significance = Mismatches = #checksum 590 5.77 114 0

Results file sqlspt.res made by on Tue 11 Apr 95 9:56:26-PDT.

Query sequence being compared:US-08-223-263-1
Number of sequences searched:
Number of scores above cutoff: 1 (1-353) 40292

Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : Swiss-Prot 30, all entries

10000-1000-5000-500-

Tue Apr 11 13:27:37 1995

Page 41

SCORE 0| STDEV -2 **೧೯೧**೫೯೦ 50---¹== 17 | 6 19 8-22-25

PARAMETERS

Times: CPU Total Elapsed 00:01:20.11 00:01:22.00	Scores: Mean Median Standard Deviation 6 8 2.07	SEARCH STATISTICS	Initial scores to save 45 Alignments to save 15 Optimized scores to save 0 Display context 10	Similarity matrix Unitary K-tuple Mismatch penalty 1 Joining penalty 2 Gap penalty 1.00 Window size Gap size penalty 0.05 Cutoff score 2 Randomization group 0
ed	viation		15 10	32 32

Number of residues: Number of sequences searched: Number of scores above cutoff: 14147368 40292 4111

Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

Page 42

Cut-off raised to 11.

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

41. 42.	39. 40.	38.	37.	36.	34.	ω.	32.	31.	30.		29.	28.	27.	26.	25.	24.	23.	22.	21.	20.	19.	18.	17.	16.	1	15.	14	1 1	12.	11.	10.	9.	.	7.	6.	5.	4.		ω!	N.	:		aomentes	
KFES_FELCA DPO1_BACCA	CNA2 HUMAN		- 1	CPS1 MOUSE			1	YAT2_SYNP6	GSPI_XANCP		CA24 MOUSE		VL2 BPV1	VSI1 REOVJ	TOG1 MOUSE	ATPD TOBAC		GAG MLVMO	HUP1 CHLKE	VTP3_TTV1V	ACRO HUMAN	EXTN SORVU	EPO MACEA	HPPK ECOLI			GAG MSVMO	OCHT YEAST			EPO MOUSE	EPO RAT	FORM MOUSE	FOR4 MOUSE		PYC1 YEAST	FHUB ECOLI	1		EPO CANFA	PLO LEPON		HOE NAME	
PROTO-ONCOGENE TYROSINE-PROTE DNA POLYMERASE I (EC 2.7.7.7)	69.7 KD PROTEIN (ORF 2). CAMP-DEPENDENT 3',5'-CYCLIC P	VATE DECARBOXYLASE (EC	NUCLEAR PORE GLYCOPROTEIN P62	CYTOCHROME P450 XXIA1 (FC 1 1	SERINEPYRUVATE AMINOTRANSFE	EXTENSIN PRECURSOR.		ATP SYNTHASE SUBUNITS REGION		**** 5 standard deviations above mean		ARMADILLO SEGMENT POLARITY PR		CURSOR	PROTEIN	_	S.	GAG POLYPROTEIN (CONTAINS: CO	H(+)/HEXOSE COTRANSPORTER.	•	ACROSIN PRECURSOR (EC 3.4.21.	EXTENSIN PRECURSOR (PROLINE-R	ERYTHROPOIETIN PRECURSOR.	2-AMTNO-4-HYDROXY-6-HYDROXYME 15	**** A grandard deviations a	BETA-MANNANASE / ENDOGLICANAS	GAG POLYPROTEIN B65 (CONTAINS	6-MANNOSYT	POLYPROTEIN (CONTAINS:			ERYTHROPOIETIN PRECURSOR.		•		PYRUVATE CARBOXYLASE 1 (EC 6.	FERRICHROME TRANSPORT PROTEIN		GLANDULAR KALLIKREIN K22 PREC	ROPOTETIN PRECIREOR (FRA	**** 8 standard deviations at	**** 9 standard deviations at	Description	
820 877	686	564	522	4.44 8 0 7 -	414	306	214	208	138	ove mean	1707	813	469	462	350	248	3010	538	533	474	421	283	192	158		1331	85.5	480	468	235	192	192	1468	1206	1180	1178	659	ove mean	259	175		above mean	Length Sc	
18 18	18	18	18	1 K	18	18	18	18	18	***	19	19	19	19	19	19	20	20	20	20	20	20	20	20	* ! * ! *	2;	5 5	2 5	2 :	21	21	21	22	22	22	22	22	***	23	24	* \ * \ *) * n * *	score sc	
71 60	62 62	43	გ :	7 5	5 8	42	43	36	36		73	67	67	63	60	36	68	S	60	61	68	4.6	4	u u	í	3 6	y (2 () بر 6 ک	ب ا	46	44	67	67	83	83	59		46	41	ŗ	3	score	•
5.80 5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80		6.29	'n	N	'n			6.77	6.77	6.77		6.77		6.77	6 77		7 25	٠,	٠;		· .			7.74	7.74	7.74	7.74	7.74		8.22	8 70	9.19		sig. Fi	
00	00	0	0	.	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0 0	0 (0 0	5	•	> 0	> 0	> c	> (0 (0	0	0	0	0	0	0		0	>	c	•	Frame	

1

43. 44.

POLG YEFV2
POLG YEFV1
AMEG_BOVIN

GENOME POLYPROTEIN GENOME POLYPROTEIN AMELOGENIN.

(CONTAINS:

3411 3411 170

18 18

48 48 36

5.80 5.32

000

US-08-223-263-1 (1-353) EPO_FELCA ERYTHROPOIETIN

PRECURSOR

(FRAGMENT).

188 AA.

STANDARD;

Tue Apr 11 13:27:37 1995

Page

43

```
Gaps
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SOFTITION DRAWN STATES OF 
                                                                                                                                                                                                                                                                                                                                                                                  MELTELLIVVMLLLTARLTLSSP---APP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTFVLLPAVDF
                                                                                                                                                                                                                                                                             70 80 90 100 110 120 1;
SLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSL--
                                                                                                                                                                  140
                                                                                                                                                                                                             NFYTWKRMDVGQQAVEVWQGLALLSE—AILRGQALLANSSQPSETLQLHVDKAVSSLRSLTSLLRALGAQ
70 80 90 100 110 120 130
                                                                                                                                                                                                                                                                                                                                               ECPALILLLSLLL---LPIGLPVLGAPPRLICDSRVLERYILEAREAENVTMGCAEGCSFSENITVPDTKV X 10 20 30 40 50 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        01-FEB-1994
01-FEB-1994
01-FEB-1994
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CARBOHYD
CARBOHYD
CARBOHYD
SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          EPO FELCA
P33708;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  DISULFID
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      GOODMAN M., BUNN H.F.;
BLOOD 82:1507-1516(1993).
BLOOD 82:1507-1516(1993).
-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN
REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE
PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MANMA
AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              CHAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      SIGNAL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           NON TER
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        PROSITE;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             EMBL;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    WEN D.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         93372347
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 SEQUENCE FROM N.A
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         EUTHERIA; CARNIVORA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                EUKARYOTA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   FELIS CATUS (CAT).
EUKARYOTA; METAZOA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                EPO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ERYTHROPOIETIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ERYTHROCYTE MATURATION;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     SUBCELLULAR LOCATION: SECRETED
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             L10606; FCERYTHRO.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 BOISSEL J.P.R.,
                                                                                                                                                                  150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          (REL.
(REL.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             188
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                1
<1
23
29
51
46
60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ₽,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  PRECURSOR (FRAGMENT).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   25%
25%
26
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          28,
                                                                                                                                                                  160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                188
183
183
55
46
60
A. 17 ...
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             20608
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Optimized Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CREATED)
LAST SEQUENCE UPDATE)
LAST ANNOTATION UPDATE)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 GLYCOPROTEIN;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    TRACY T.E.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             WW.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           BY SIMILARITY.
ERYTHROPOLETIN.
BY SIMILARITY.
BY SIMILARITY.
POTENTIAL.
POTENTIAL.
POTENTIAL.
                                                                                                                                                                  170
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             175419
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    MULCAHY L.S.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Q.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 HORMONE;
                                                                                                                                                                  180
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           533
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   SIGNAL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Significance
Mismatches
                                                                                                                                                                     190
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CZELUSNIAK
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ADULT MAMMALS
                                                                                                                                                                     200
                                                                                                                                                                                                                                                                                                       130
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      J
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        0 0 0
                                                                                                                                                                                                                                                                               LGTQLPPQG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                9.19
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     128
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            3HI
```

Initial Residue

Score Identity

H B B

24 28% 32

Optimized Score Matches

53 53

Significance Mismatches

0 11

.70 98 0

8

Conservative Substitutions

5

20

CARBOHYD CARBOHYD NON TER SEQUENCE

AA;

19193

Œ,

155009

Ω II

POTENTIAL. POTENTIAL.

DISULFID DISULFID CARBOHYD

<1 23 29 51 46 60 105 175

22 >175 >175 >175 105 175

ERYTHROPOIETIN. BY SIMILARITY. BY SIMILARITY. POTENTIAL.

BY SIMILARITY

NON TER

CHAIN

MELTELLLVVMLLLTARLTLSSP---APP--ACDLRVL---SKLLRDSHVLHSRLSQCPEVHPLPTPVLLPA

30

40

5

60

120 ---SGQVRL--LLGAL-

Listing for Mary Ha O

1

Tue Apr 11 13:27:37

```
210
SA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        US-08-223-263-1 (1-353)
EPO_CANFA ERYTHROPOIETIN PRECURSOR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      -KEATSLPEA--TSAAPLRTFTV
140 150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          RTTAHKDPNAIFLSFOHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTA
                                                                                                                                                                                                                                                                                                                                                                                              EPO CANFA
P33707;
01-FEB-1994 (
01-FEB-1994 (
01-FEB-1994 (
                                                                                                                                                                  WEN D., BOISSEL J.P.R., TRACY T.E., MULCAHY L.S., CZELUSNIAK J., GGODMAN M., BUNN H.F.;
BLOOD 82:1507-1516(1993).
-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF PHYSIOLOGICAL LEVEL OF CHECOLATING ERYTHROCYTE MASS.
-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMA.
AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
-!- SUBCELLULAR LOCATION: SECRETED.
                                                                                                                                                                                                                                                                                                                                                                                 01-FEB-1994 (REL. 28, CRE 01-FEB-1994 (REL. 28, LAS 01-FEB-1994 (REL. 28, LAS ERYTHROPOIETIN PRECURSOR
                                                                                                                                                                                                                                                                                                                                          CANIS FAMILIARIS (DOG).
EUKARYOTA; METAZOA; CHORDATA;
                                                                                                                                                          EMBL;
                                                                                                                                                                                                                                                                                        SEQUENCE FROM N.A.
93372347
                                                                                                                                                                                                                                                                                                                              EUTHERIA;
                                                                                                                                ERYTHROCYTE MATURATION;
                                                                                                                                              PROSITE;
                                                                                                                                            L13027; CFERYPRE. 
E; PS00817; EPO.
                                                                                                                                                                                                                                                                                                                               CARNIVORA
                                                                                                                                                                                                                                                                                                                                                                                                                                                  STANDARD;
                                                                                                                                                                                                                                                                                                                                                                                              CREATED)
LAST SEQUENCE UPDATE)
LAST ANNOTATION UPDAT
                                                                                                                               GLYCOPROTEIN; HORMONE;
                                                                                                                                                                                                                                                                                                                                                                                  (FRAGMENT)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      -DTIC--KLFRIYSNFLRGKLTLYTGEACRRGDR
                                                                                                                                                                                                                                                                                                                                          VERTEBRATA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                  PRT;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           (FRAGMENT)
                                                                                                                                                                                                                                                                                                                                                                                                                                                  175 AA.
                                                                                                                                                                                                                                                                                                                                            TETRAPODA; MAMMALIA;
                                                                                                                                SIGNAL
                                                                                                                                                                                               ADULT MAMMALS
                                                                                                                                                                                                                         SHE
```



Ŋ.

4

3

Tue Apr 11 13:27:37 1995

Page

45

US-08-223-263-1 (1-353)
KIKL_MOUSE GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35)

01-APR-1990 (REL. 14, CREATED)
01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
GLANDULAR KALLIKELE 18, K22 PRECURSOR (EC 3.4.21.35) (TISSUE (MGK-22) (EPIDERMAL GROWTH FACTOR-BINDING PROTEIN TYPE A)
(NERVE GROWTH FACTOR BETA CHAIN ENDOPEPTIDASE) (BETA-NGF-KLKL MOUSE P15948; ENDOPEPTIDASE). STANDARD; PRT; 259 B KALLIKREIN) (EGF-BP A)

MUS MUSCULUS (MOUSE).
EUKARYOTA; METAZOA; C
EUTHERIA; RODENTIA. CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

DRINKWATER C.C., EVANS B.A., RICHARDS BIOCHEMISTRY 26:6750-6756(1987). 88107594

STRAIN=BALB/C; SEQUENCE FROM N.A.

TISSUE=SALIVARY GLAND;

FAHNESTOCK M., WOO J.E., MOBLEY W.C.; SEQUENCE OF 25-54 LOPEZ G.A.,

BIOCHEMISTRY 30:3443-3450(1991). SEQUENCE OF 25-41 SNOW J., WALZ D.A., ARICI M.J.,

MIYAKE

TISSUE-SUBMANDIBULAR GLAND; PETERS J., TAKAHASHI S., TADA M., J. BIOCHEM. 111:643-648(1992). SEQUENCE OF 17-54 AND 70-120 FROM z

EVANS

EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;

J. BIOL. CHEM. 262:8027-8034 (1987).

-i- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS IN KININOGEN TO RELEASE LYS-BRADYKININ.

-i- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF MET-|-XAA OR LEU-|-XAA.

MET-|-XAA OR LEU-|-XAA.

-i- SIMILARITY: TO OTHER TRYPSIN-LIKE SERINE PROTEASES. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.

EMBL; M17977; MMGFBPE1.

EMBL; M17979; MMGFBPE3. ÷

+

A S No.

- 3

Yes.

1

À. r Min

Listing for

Mary Hale

Tue Apr 11, 13:27:37 1995

Page

46

```
Residue Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               HPDFNMSLLOSVPTGADLSNDLMLLRLSKPADITDVV--KPIDLPTTEPKLGSTCLASGMGSINQLIYQ---
100 110 120 130 140 150 160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            KNSQPWQVAVYYLDEYLCGGVLLDRNWVLTA-
40 X 50 60
                                                                                                                                                                       NPNDIQCVSIKLHPNEVCVKAHILKVTDVMLCAGEMNGGKDTCKGDSGGPLIC-
170 180 190 200 210
                                                                                                                                                                                                                                                   LLGALQSLLGTQLPPQGRTTAH-KDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLI
                                                                                                                                                                                                                                                                                                                                                                                                                       60 100 110 HPIPTPVLLPAV--DFSIGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLG-
                                PCGEPN--
                                                                            200 210 220 230 240 LNELPHRTSGLLETNFTASA-RTTGSGLLKWQQGFRAKIPGLLNQTSRSL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CHAIN
ACT SITE
ACT SITE
ACT SITE
DISULFID
DISULFID
DISULFID
DISULFID
DISULFID
DISULFID
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         EMBL; M18618; MMKALF
PIR; A29746; A29746.
PIR; A38356; A38356.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          CARBOHYD
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             PROPEP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        SIGNAL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  MULTIGENE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               HYDROLASE;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       PROSITE; PSO0134; TRYPSIN_HIS.
PROSITE; PSO0135; TRYPSIN_SER.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    M18598;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              X 10
MELTELLIVVMLL----LTARUTLSSPAPPACD--LRVLSK--LLRDSHVLHSRLSQCPEV
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       E FAMILY; ZYMOGEN; SIGNAL.

1 17

1 27

1 27

1 37

1 4

ACTIVATION PEPT 18

2 5 259

CHANGE RELAY SY 118

1 18

CHARGE RELAY SY 211

2 11

2 11

2 11

2 11

3 171

3 171

3 171

3 171

3 171

3 171

3 171

3 171

3 171

3 171

3 171

3 171

4 18 CHARGE RELAY SY 211

3 171

3 171

3 171

3 171

3 171

3 171

3 171

4 18

5 196

6 3 19 SIMILARITY.

5 197

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 102

1 103

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 104

1 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  FAMILY;
-APAIYTKLIKFTSWIKDTMAKNP
240 250 X
                                                                                                                                                                                                                                                                                      140
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                23
23
45
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Optimized Score
Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Conservative Substitutions
                                                                                                                                                                                                                                                                                      150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CHARGE RELAY SYSTEM.
BY SIMILARITY.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     ACTIVATION PEPTIDE.
GLANDULAR KALLIKREIN I
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
                                                                                                                                                                                                                                                                                   160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            -AHCYEDKYNIWLGKNKLFQDEPSAQHRLVSKSFP
70 80 90
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        (I
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             GLYCOPROTEIN;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             59
59
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Significance
Mismatches
                                                                                                                                                                    -DGVLQGITSWGST
220 230
                                                                                                                                                                                                                                                                                   80
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   0 0
                                                                                                                                                                                                                                                                                                                                                                                                                          QLSGQVRI
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                8.22
142
0
```

120

B35586 US-08-223-263-1 (1-353) FHUB_ECOLI FERRICHROME TRANSPORT PROTEIN FHUB 01-APR-1988 01-OCT-1994 FERRICHROME FHUB ECOLI P06972; 01-APR-1988 (REL. 07, CREATED)
(REL. 07, LAST SEQUENCE UPDATE)
(REL. 30, LAST ANNOTATION UPDATE)
TRANSPORT PROTEIN FHUB PRECURSOR. STANDARD; PRT; 659 PRECURSOR



FHUB

ENTEROBACTERIACEAE ESCHERICHIA COLI.
PROKARYOTA; GRACILICUTES;

SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

Tue Apr 11 13:27:37 1995

Page 47

```
Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            SO TITIES DE RECOCCOCCOCCOCCER REPORT RESERVATION DE RECOCCOCCOCCER RESERVATION DE RESERVATION D
180
                                                                                                                                                                                                                                                                                                                                                                        RDAHGWTWASGALLEDIMPWRWPRIMAALFAGVML---ÀVAGCIIQRLTGNPMASPEVLGISSGAAFGVVIM 370 380 390 400 410 420 430
                                            LLMMIQ-ASGDPRMAQVITWISGSTYNATDAQVWRTGIVMVILLAITPICRRWITIILPI-GGDT---ARAVG
500 510 520 530 540
                                                                                                                               LLGALQSLLGTQLPPQ-
                                                                                                                                                                                                                                                                                             LPTPVLLPAVDF--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  SIGNAL
CHAIN
CONFLICT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                HOL.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ++
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  EMBL;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    BURKHARDT R.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      SEQUENCE OF 1-23 FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        FUJITA N., MORI H., YURA T., ISHIHAMA A.; NUCLEIC ACIDS RES. 22:1637-1639(1994).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               94261430
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     STRAIN=K12 / W3110;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   KOESTER W., BRAUN V.;
MOL. GEN. GENET. 204:435-442(1986)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  RON TRANSPORT;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         Identity
                                                                                                                                                                                                                                        -FLVPGNAFGWLLPAGSLG---
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           DICITRATE TRANSPORT SYSTEM.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED. SIMILARITY: TO THE FECC AND FECD PROTEINS OF THE IRON
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              GEN. GENET. 209:49-55(1987).
FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN AFFINITY TRANSPORT OF IRON(III)-FERRICHROME INTO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 X04319; ECFHUB.
; D26562; EC82K.
; X05810; ECFHUACD.
S07318; S07318.
                                                                                                                                                           130
190
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  EG10303; FHUB.
ANSPORT; TRANSPORT; INNER MEMBRANE; SIGNAL.
1 20
1 20
21 659 FERRICHROME-IRON TRAN
21 659 I -> V (IN REF. 3).
                                                                                                                                                                                                                                                                                                                                                                                                                                                     MELTELL----LVVMLLLTARLTLSSPAPPACDLRVLSKLLRDS-HVL-HSRLSQCPEVHP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    BRAUN V.;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Ą,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     22
24%
77
200
                                                                                                                                                                                                                                                                                       70 100 110 120
SLGEWKTQMEETKAQDILGAVILLLEGVMAARGQLGPTCLSSLLGQLSGQVRL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            70335 MW;
                                                                                                                                  GRTTAHKD-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Optimized Score = 59
Matches = 79
Conservative Substitutions
                                                                                                                                                           140
210
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    L -> V (IN KE); 2233995 CN;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     FERRICHROME-IRON TRANSPORT L -> V (IN REF. 3).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 20
                                                                                                                                                                                                               -AAVTLLIIMIAAGRGGFSPHRMLLAGMALSTAFTM
460 470 480 490
                                                                                                                            -PNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPP
220
                                                                                                                                                           150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             į
230
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Significance
Mismatches
                                                                                                                                                           160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                N
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 40
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         HE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                THE HIGH-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              PROTEIN FHUB
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       E.COLI
240
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         (III)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          7
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     7.74
170
0
250
```

Listing ğ Mary Hale Tue Apr 11 13:27:37 1995

8

```
US-08-223-263-1 (1-353)
PYC1_YEAST PYRUVATE C
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SALVGGLLLVFADWCGRMVLFPFQIPAGLLSTFIGAPYFIYLLRKQSR
620 630 640 650 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     260 270 280 290 X 300 HELLNGTRGLF-PGPSRRTLGAPDI8SGTSDTGSLFP---NLQPGYSPSPTHFPTGQY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    MALTPTRIAL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               TTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRI
                                                                                                                                EMBL; J03889; SCPCB.
PIR; A2923; OYBYP.
PROSITE; PS00188; BIOTIN.
PROSITE; PS00866; CPSASE 1.
PROSITE; PS00867; CPSASE 2.
                                                                                                                                                                                                                                                                                                  MORRIS C.P., LIM F., WALLACE J.C.;
BIOCHEM. BIOPHYS. RES. COMMUN. 145:390-396(1987).
-i- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
-i- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +
ORTHOPHOSHATE + OXALOACETATE.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      PYC1 YEAST
P11154;
                                                           ACT SITE
BINDING
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         01-UUL-1989 (REL. 11, CREATED)
01-UUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
01-UCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
PYRUVATE CARBOXYLASE 1 (EC 6.4.1.1) (PYRUVIC
              SIMILAR
SIMILAR
SIMILAR
                                                                                                                                                                                             -!- PATHWAY: GLUCONEOGENESIS.
-!- SUBUNIT: HOMOTETRAMER.
-!- COFACTOR: BIOTIN, AND ZINC.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- SUBCELLULAR TOCATION: CYTOPLASMIC.
-!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSERRASES AND CARBAMYL PHOSPHATE SYNTHETASES.
EMBL; J03889; SCECB.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               PYC1 OR PYV.
                                                                                                                                                                                                                                                                                                                                                                                                                            87241529
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      I MIT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SACCHAROMYCES CEREVISIAE (BAKER'S YEAST)
EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCO
                                                                                          NP BIND
                                                                                                        MULTIGEŅE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               SEQUENCE FROM N.A., AND PARTIAL SEQUENCE
                                                                                                                        LIGASE;
                                                                                                                                                                                                                                                                                                                                                                                                                                          SEQUENCE OF 1003-1178 FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                          ^{\sim}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        BIOL.
                                                                                                                                                                                                                                                                                                                                                                                                                                                               MORRIS C.P., OCCHIODORO F., N
. CHEM. 263:11493-11497(1988).
                                                                                                                        MULTIFUNCTIONAL
                               182
312
1135
160
350
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     STANDARD;
               187
312
1135
1330
470
1178
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CARBOXYLASE 1 (EC 6.4.1.1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      570
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  -LILAACLTATATMTIGPL--SFVGLMAPHIARMMGFRRTMPHI--
                                                                                                                      ENZYME;
BY SIMILARITY.

BIOTIN (BY SIMILARITY).

CARBAMOYL PHOSPHATE SYNTHETASES.

WITH OTHER BIOTIN CARBOXYLASES.

WITH OTHER BIOTIN CARRIER PROTEINS

WITH LIPOAMIDE ACETYLTRANSFERASE.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      580
                                                                                       ATP (POTENTIAL).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     PRT;
                                                                                                                     BIOTIN; GLUCONEOGENESIS; ZINC;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                HEMIASCOMYCETES
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     1178 AA.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      WALLACE J.C.;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      590
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            CARBOXYLASE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   (PYRUVIC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 CARBO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ۳
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            (PCB
                 AND
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          <u>:</u>
```



100

Tue Apr 11 13:27:37 1995 Page 49

Initial Score =
Residue Identity =
Gaps = SS 630 LFPLPPTLPTPVVQLHPLLPDPSAPTPTPTSPLLNTSYTHSQNL--SQEG KIKSNPSVPHLHDAQGNVINVTKSAPPSGWRQVILIEKGPAEFAR--510 520 530 540 60 70 110
80 90 100 110
TPVILPAVDFSLGEWKTQMEETKAQDIIGAVTILLIEGVMA--ARGQLGPTCLSSLIGQUSGQVR--RIRGVKTNIPFLLTLLTNPVFIEGTYWGT-FIDDTPQLFQMVSSQNRAQKLLHYLADVADNGSSIKGQIGLP 440 X 450 460 470 480 490 500 SGDMLQPGKKYNLDYYLEIAEKIVQMG--THILGIKD-MAGTMKPAAAKLLIGSLRAKYPDLPIHVHT--710 720 730 740 750 760 PN I PFQMLLRGANGVAYSSLPDNA I DHFVKQAKDNSVD I FRVFDALNDLEQLKVGVDAVKKAGGVVEATVCF PSRTSLVLTLNELPNRTSGL---LETNFTASA---DTTWRDAHQSLLATRVRTHDLATIAPTTAHALAGAFALECWGGATFDVAMRFLHEDPWQRLRKLR--HDSAGTRVASMTACALAGADVVDVAINSMSGL--TSQPSINALLASLEGNIDTGINVEH 770 780 790 800 810 X 820 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
PYRUVATE CARBOXYLASE 2 (EC 6.4.1.1) (FYRUVIC
PYC2 OR YBR218C OR YBR1507. PYC2 YEAST STANDARD; P32327; 01-0CT-1993 (REL. 27, CREATED) STUCKA R., DEQUIN S., SALMON J.-M., GANCEDO C.; MOL. GEN. GENET. 229:307-315(1991). SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES. SEQUENCE SEQUENCE FROM N.A., AND CHARACTERIZATION --LGALQSLLGT---640 X 10 20 30 MELTELLLVVMLLLTARLTLSSPAPPACDLRVLS----1178 AA; 140 150 160 170 180 ---QLPPQGRTTAHKDPNAIFLSFQHLLRGKV---RFLMLVGGSTLCVRRAPPTTAV 22 23% 79 650 130099 MW; Conservative Substitutions Matches Optimized Score = 590 210 660 PRT; 1180 AA. 7059028 600 670 -RTTGSGLLKWQ--QGFRA--KIPGLLNQT-SR Ω, 68 92 CARBOXYLASE 610 (PYRUVIC CARBO -KLLRDSHVLHSRLSQCPEVHPLP Significance = Mismatches = = 2) ---QVRQFNGTLLM 550 560 690 230 (PCB 50

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

Page 50

```
Initial Score
Residue Identity
                                                                                                                         RIRGVKTNIPFILTLINPVFIEGT-YWTTFIDDTPQLFQMVSSQNRAQKLLHYIADIAVNGSSIKGQIGLF
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              [2]
SEQUENCE FROM N.A.
STRAIN=S288C;
EL BAKK
60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   SCHERENS B., VIERENDEELS F.;
SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
-I- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
-I- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +
                                                                                                                       CONFLICT
CONFLICT
SEQUENCE
                                                                                                                                                                        CONFLICT
CONFLICT
CONFLICT
                                                                                                                                                                                                      CONFLICT
CONFLICT
CONFLICT
                                                                                                                                                                                                                                                                                                                                                                                       EMBL; X59890; SCPYC2G.
EMBL; Z36087; SCYBR218C.
PIR; S46094; S46094.
                                                                                                                                                                                                                                                                                                                                                                                                                                       -!- PATHWAY: GLUCONEOGENESIS.
-!- SUBUNIT: HOMOTETRAMER.
-!- COFACTOR: BIOTIL, AND ZINC.
-!- SUBCELLUIAR LOCATION: CYTOPLASMIC.
-!- INDUCTION: BY GLUCOSE.
                                                                                                                                                                                                                                                                                                                    NP BIND
                    440
                                                                                                                                                                                                                                                                                                                                                            PROSITE; PS00188; BIOTIN.
PROSITE; PS00866; CPSASE 1.
PROSITE; PS00867; CPSASE 2.
                                                                                                                                                                                                                                                        CONFLICT
                                                                                                                                                                                                                                                                                                          BINDING
                                                                                                                                                                                                                                                                                                                                         MULTIGENE FAMILY.
                                                                                                                                                                                                                                                                                                                                                           PROSITE;
                                                                                                                                                                                                                                                                                                                                                                                                                     -!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.
                                                                                                                                                      CONFLICT
                                                                                                                                                                CONFLICT
                                                                                                                                                                                                                                               CONFLICT
                                                                                                                                                                                                                                                                             SIMILAR
                                                                                                                                                                                                                                                                                       SIMILAR
                                                                                                                                                                                                                                                                                                SIMILAR
                                                                                                                                                                                                                                                                                                                                                IGASE;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ORTHOPHOSPHATE + OXALOACETATE.
                    X 450
                                                                                                                                                                                                                                                                                                                                                 MULTIFUNCTIONAL
                                                MELTELLLVVMLLLTARLTLSSPAPPACDLRVLS--
70
                                                                                                                        1180
1180
                                                                                                                                            1001
1155
1178
                                                                              23#
65
                                                                                                                       AA;
                                                                                                                                            831
839
1001
1155
1178
                                                                                                                                 1180
                                                                                                                                                                                                                                                                            1180
 80
                    460
                                                                              Conservative Substitutions
                                                                                        Optimized Score
Matches
                                                                                                                       130166
                                                                                                                                                                                                                                                                                                                                                  ENZYME; BIOTIN; GLUCONEOGENESIS;
                                                                                                                                                                                 GTA -
                                                                                                                                                                                                                                    WITH
                                                                                                                                                                                                                                                                           CARBAMOYL PHOSPHATE SYNTHETASES. WITH OTHER BIOTIN CARBOXYLASES. WITH OTHER BIOTIN CARRIER PROTEINS AND
                                                                                                                                                                                                                                                                                                        BY SIMILARITY.
BIOTIN (BY SIMILARITY).
90
                                                                                                                                                                                                                                                                                                                             ATP (POTENTIAL).
                                                                                                                                                                                                       , ,
                                                                                                                                  ¦
                                                                                                                                                                          Ļ
                                                                                                                                                                                                                           Ŷ
                    470
                                                                                                                                                                                H LIPOAMIDE ACETYLTRANSFERASE.

C (IN REF. 1).

E (IN REF. 1).

K (IN REF. 1).

(IN REF. 1).

IN REF. 1).

T (IN REF. 1).

T (IN REF. 1).

T (IN REF. 1).

R (IN REF. 1).

R (IN REF. 1).
                                                                                                                                            ש
                                                                                                                                  KVIFTR
100
                                                                                                                                            REF
                                                                                                                                                                          REF. 1)
                                                                                                                                                              REF
                                                                                                                         S
                    480
                                                                                         91
91
                                                -KLLRDSHVLHSRLSQCPEVHPLP
                                                                                        Significance = Mismatches =
110
                                                                                                                                 <u>.</u>
                    490
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             PIERARD A.,
                                                                                                                                                                                                                                                                                                                                                  ZINC;
120
                                                                               11 11
                    500
                                                           50
                                                                             7.74
237
0
```

51

Page

US-08-223-263-1 (1-353) FOR4_MOUSE FORMIN 4. MOUSE

FOR4 MOUSE STANDARD; PRT; 1206 AA. Q05839; Q0-JUN-1994 (REL. 29, CREATED) Q1-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE) Q1-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE) FORMIN 4. MUS MUSCULUS (MOUSE). EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; EUTHERIA; RODENTIA. TETRAPODA; MAMMALIA;

SEQUENCE FROM N.A.

GENES DEV. 6:29-37 (1992).

GENES DEV. 6:29-37 (1992).

-I- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB
-I- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB
AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR
BE INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
-I- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
HAS A FUNCTIONAL ROLE ONLY IN THE LIVER AND LIMB.
-I- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL

<u>!</u> LIMB BUD.

ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT ę THE DEVELOPING

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

Page

52

Residue Identity Initial US-08-223-263-1 (1-353) FORM MOUSE FORMIN. EMBL; X62377; PRICE CONTROL OF THE STATE OF FELFETLEVV--QLHPLLPDPSAPTFTFTS-GYLNRIHELLNGTRGLFPGPSRRTLGAP~DISSGTS----DTGSLP--RLQAVWPPPKTKDTEEKVGLKYTEAEYQAAI----LHLKREHKEEIETLQAQFELKTFHIRGEHALVTAR--480. 490 500 510 520 530 EQLSQLINIDMPRTEQKEADPEFHGADEMGYSTDQESHKSP-RDAHVQGGQVKARTPETALEAFKALFIRP-350 X 360 370 380 390 400 410 LVLTLNE--190 ----Leeaienlkqqlekrregceemrdvcistdddcspkafrnvciqtdretfikpcdaeskatrs-sqi-540 550 560 570 580 590 600 EMBL; X62379; MMIFOR. PIR; S24407; S24407. --VPKKLTISLTQLSPSKDSKDIHAPFQTREGTSSSSQQKISPPAPPTPPPLPPPLIPPPPLPPG---L 610 620 630 640 650 660 -PKKGSTADTSELEALKRKKHEK--ESLRAV----FERSKSRPADSPSDPKSPDQSPTEQDDRTPG 420 430 440 450 460 470 DETERMINED TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR 130 140 150 160 170 180 --SILGTQLPPQGRTTAHKDPNAIFLSFQHLLR-GKVRFIMLVGGSTLCVRRAPPTTA-VPSRTS X 10 20 30 40 50 MELTE-----LLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVL--HSRLSQCPE 8 A B 260 -LPNRTSGLLE-TNFTASARTTGS----GLLKWQQGFR---AKIPGLLNQTSRSLDQIP STANDARD; 22 23**%** 95 Optimized Score = Matches = Conservative Substitutions 1468 AA. --PLLNTSYTHSQNLSQEG ALTERNATIVE SPLICING 220 Ş 67 95 Significance = Mismatches = PNLQPGYSPSPTHPPTGQYTI 350 7.74 220 0

PARAG FORM MOUSE Q05860; Q1-JUN-1994 Q1-JUN-1994 Q1-JUN-1994 (REL. 29, CREATED)
(REL. 29, LAST SEQUENCE UPDATE)
(REL. 29, LAST ANNOTATION UPDATE)

Listing for Mary Hale

1 C ...

Tue Apr 11 13:27:37 1995

Page 22

```
Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       CC -I- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY CC AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY HE CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.

CC -I- SUBCELLULAR LOCATION: NUCLEAR.

CC -I- SUBCELLULAR LOCATION: NUCLEAR.

CC -I- SUBCELLULAR LOCATION: NUCLEAR.

CC -I- ISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS, CLIMB, OVARY, BRAIN, SMALL INTESTINE, SALLVARY GLAND AND HARDERIAN CLIMB, OVARY, BRAIN, SMALL INTESTINE, SALLVARY GLAND AND HARDERIAN CLIMB, OVARY, BRAIN, SMALL INTESTINE, SALLVARY GLAND AND HARDERIAN CLIMB, OVARY, BRAIN, SMALL INTESTINE, SALLVARY GLAND AND HARDERIAN CLIMB, OVARIAN AND THE APICAL ECTODERWAL RIDGE AND THE PROPERING COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING KINNEY MORPHOSENESIS, EXPRESSION IS INITIALLY RESTRICTED TO CLIMBARY MORPHOSENESIS, EXCRESSION INITIALLY RESTRICTED TO CLIMBARY MORPHOSENESIS.
                                                                                                                                                                                                                                                                                                                                                 SADTLEPSSTTKVTETKGASPTSLRASQTWLVSEEASEKGLGPEKITAPPQHQLPPGIASEGFPC-DNFKEQ
170 580 590 600 610 620 630
  ΝË-
                                                                      WPPPKTKDTEEKVGLKYTEAEYQAAI---
710 720
                                                                                                                                                                                                         TAKDLPNKD--GGVWVPGYRAGPPCPFL----LHEEKEKTSRSEL
0 650 660 670
                                                                                                                                                                                                                                                                             TPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQ-LSGQVRLLLGALQ--
                                                                                                                                                                                                                                                                                                        60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       PIR; S11
NUCLEAR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          MUS MUSCULUS (MOUSE).
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
EUTHERIA; RODENTIA.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  WOYCHIK R.P., MAAS R.L., NATURE 346:850-853(1990)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     90363291
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   TISSUE=KIDNEY, AND TESTIS;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Score
                                                                                                                                     -SILGTQLPPQGRTTAHKDPNAIFLSFQHLLR-GKVRFIMLVGGSTLCVRRAPPTTA-VPSRTSIVLTL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Identity
PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING
298 301 POLY-SER.
861 864 POLY-SER.
869 270 PRO-RICH.
                                                                                                                                                                                                                                                                                                                                                                                                                     X 10 20 30 40 50 MELTELLLYVMLLITARLT-LSSPAPPACDLRVLSKLLRDSHVL----HSRLSQCPEVHPLP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       H
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     1468 AA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                861
869
977
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           22
23%
86
                                                                                                                                                                                                                                                                                                        80
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Optimized Score =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       163809 MW;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ZELLER R.,
                                                                 --LHLKREHKEEIETLQAQFELKTFHIRGEHALVTAR---
730 740 750 760
                                                                                                                                                                                                                                                                                                        90
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        POLY-SER.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       11493196
                                                                                                                                                             160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          VOGT T.F.,
                                                                                                                                                                                                                                                                                                        100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       2
                                                                                                                                                             170
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   67
95
                                                                                                                                                                                                         --YLDLNPDQSPTEQDDRTPGRLQAV
680 690 700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             LEDER
                                                                                                                                                                                                                                                                                                        110
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Mismatches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Significance
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ۳.
                                                                                                                                                             180
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            TESTIS,
ND HARDERIAN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   0 0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             7.74
224
0
                                                                                                                                                                  190
```

Residue Initial SOUTH THE TENTON OF THE TENTON US-08-223-263-1 (1-353)
EPO_RAT ERYTHROPOIETIN PRECURSOR EPO RAT STANDARD; PKT; 132 cm.
PZ9676;
O1-APR-1993 (REL. 25, CREATED)
O1-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
O1-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE) CARBOHYD CARBOHYD BIOCHIM. BIOPHYS. ACTA 1171:99-102(1992).

-!- FUNCTION: ERYTHROPOLETIN IS THE FRICIPAL HORMONE INVOLVED IN THE REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.

-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS AND BY LIVER OR FETAL OR NEONATAL MAMMALS.

-!- SUBCELLULAR LOCATION: SECRETED. DISULFID CARBOHYD EMBL; D10763; RNEPO. PIR; S28148; S28148. PROSITE; PS00817; EPO. NAGAO M., SUGA H., SASAKI R.; RATTUS NORVEGICUS (RAT) EUKARYOTA; METAZOA; CHO EPO. SEQUENCE CHAIN SIGNAL ERYTHROCYTE 9304201 STRAIN=WISTAR; TISSUE=KIDNEY; EUTHERIA; SEQUENCE FROM N.A. × Identity 10 E MATURATION;
1 26
1 192
27 192
33 187
50 50
64 64
109 109 RODENTIA. H H H METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; 1 27 33 50 64 109 192 AA; 21 27**%** 24 OKANO M., MASUDA S., 21286 Optimized Score Matches Conservative Substitutions 4; GLYCOPROTEIN; HORMONE; S
5 BY SIMILARITY.
5 ERYTHROPOLETIN.
7 BY SIMILARITY.
9 BY SIMILARITY.
9 BY SIMILARITY.
9 BY SIMILARITY.
1286 MW; 179242 CN; 30 NARITA 40 44 53 н., SIGNAL. Significance Mismatches IKURA 50 ... ; 0 0 0 8 7.25 118 0 OF /



23

· Sex series

W.W.

4 10

1

16

¥.

Tue Apr 11 13:27:37 1995

Page 55

10. EPO US-08-223-263-1 (1-353) MOUSE ERYTHROPOIETIN PRECURSOR

MUS MUSCULUS (MOUSE). EUKARYOTA; METAZOA; C EUTHERIA; RODENTIA. EPO MOUSE STANDARD; PRT; 192 AA. P07321; 01-APR-1988 (REL. 07, CREATED) 01-APR-1998 (REL. 07, LAST SEQUENCE UPDATE) 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE) ERYTHROPOIETIN PRECURSOR. SEQUENCE FROM N.A. 87039105 EPO. SHOEMAKER C SEQUENCE FROM N.A. C.B., MITSOCK L.D.; BIOL. 6:849-858(1986). CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

87039104

MCDONALD J.D., LIN F.-K., GOLDWASSER E.;

MOL. CELL. BIOL. 6:842-848(1986).

-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.

-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADDLT MAMMALS AND BY LIVER OF FETAL OR NEONATAL MAMMALS.

-!- SUBCELLULAR LOCATION: SECRETED.

EMBL; ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.

SIGNAL

1 26
CHAIN
27 192
ERYTHROPOIETIN.
DISULFID
33 187
BY SIMILARITY.
CARBOHYD
50 50
BY SIMILARITY.
CARBOHYD
64 64
BY SIMILARITY.
CARBOHYD
109 109
BY SIMILARITY. SEQUENCE ROSITE; L; M12482; MMERP. L; M12930; MMERPA. ; A24901; A24901. ; A24902; A24902. SITE; PS00817; EPO. ξ 21365 ¥. ERYTHROPOIETIN.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY. 188541 CN;

. V to

.

1,26

Acres 1

出於整

.- 34

+ 26

36

47

1

4

Listing for Mary Ha , a 0

Tue Apr 11 13:27:37 1995 1

Page

56

US-08-223-263-1 (1-353)
GAG_MLVAB GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; INNER

GAG.
ABELSON MURINE LEUKEMIA VIRUS.
VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE; GAG MIVAB 21-JUL-1986 (REL. 01, CREATED)
21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; CORE SHELL PROTEIN P30) STANDARD; PRT; 235 AA. INNER COAT PROTEIN P12;

SEQUENCE FROM N.A. 83221648

REDDY E.P., SMITH M.J., SRINIVASAN A.;
PROC. NATL. ACAD. SCI. U.S.A. 80:3623-3627(1983).
PROC. PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
-!- THIS PROTEIN IS SYNTHESIZED AS A GAG-ABL POLYPROTEIN.

EMBL; V01541; REAMLV. PIR; A03931; FOMVGM.

COAT PROTEIN; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION
CHAIN 2 131 CORE PROTEIN P15.

CHAIN CHAIN CHAIN LIPID 132 216 216 131 215 235 INNER COAT PROTEIN CORE SHELL PROTEIN MYRISTATE. P12

Initial Score SEQUENCE 235 AA; 25641 MW; Optimized Score 303652 CN; II

Gaps Residue Identity 170 180 190 200 210 TLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLD 11 11 11 21 23* 29 Conservative Matches Substitutions 42· Significance Mismatches II 7.25 106 0

```
Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              12.
                                                                                                                                                                                                                                                                                                                                                                SO FITTER AREA CONTROL OF THE SOURCE CONTROL
                                                                                                                    170 180 190 200 210 TLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLD
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               US-08-223-263-1 (1-353)
GAG_MSVMT GAG POLYPROTEIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            QIPGYINRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           PTPVVQLHPLLPDPSAPTPTSPLLNTS---YTHSQNLSQEG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   PROGIFNEDLITOVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPPWVKPFVHPKP---PP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          MURPHY E.C.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            GAG MSVMT STANDARD; PRT; 468 AA. p32594; CREATED) 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE) 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE) 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE) 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE) CREATED COLYPROTEIN (CONTAINS: CORE PROTEIN P15;
                                                                                                                                                                                                                                                                                                                                                                                               CHAIN
CHAIN
CHAIN
LIPID
                                                                                                                                                                                                                                                                                                                                                                                                                      EMBL; M96854; MMSAAX.
EMBL; S42321; S42321.
PIR; A42745; FOMYMU.
COAT PROTEIN; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION.
CHAIN 2 131 CORE PROTEIN P12.
CHAIN 132 215 INNER COAT PROTEIN P12.
CHAIN 216 468 CORE SHELL PROTEIN P30.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 MOLONEY MURINE SARCOMA VIRUS (STRAIN TS110).
VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CORE SHELL PROTEIN P30).

    J. VIROL. 66:5329-5337(1992).
    -!- PIM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
    -!- THIS PROTEIN IS PROBABLY TRANSLATED AS A GAG-MOS POLYPROTEIN.

                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ONCOVIRINAE.
                                                                                                                                                                                                                                                                                                                                                                SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         92365121
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     EQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                    Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ---SLPLEPPLSTPPRSSLYPALTPSLGAKPKPQVLSDSGGPLIDLLTED
120 130 140 150 160
                                                                                                                                                                                                                                                                  Identity
250
                                                     MGQTVTTPLSLTL-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       MGQTVTTPLSLTL
X 10
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          CHIOCCA S.M.,
                                                                                                                                                                                                                                               0 0 0
                                                                                                                                                                                                                                                                                                                                                                468 AA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 260
260
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             70
                                                                                                                                                                                                                                                                                                                                                                52681 MW;
                                                                                                                                                                                                                                                                     Optimized Score Matches
                                                                                                                                                                                                                                        Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          GILBRETH M.A.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 270
270
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             80
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     (CONTAINS: CORE PROTEIN P15; INNER
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             -GHWKDVERIAHNQSVDVKKRRWVTFCSAEWP
20 30 40
                                                                                    -DHWKDVERIAHNQSVDVKKRRWVTFCSAEWP
                                                                                                                                                                                                                                                                                                                                                                                            INNER COAT PROTEIN P12.
CORE SHELL PROTEIN P30.
MYRISTATE (BY SIMILARITY
                                                                                                                                                                                                                                                                                                                                                                RISTATE (BY SIMILARITY).
1077859 CN;
280
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 350 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             90
                                                                                                                                                                                                                                                                                                    II
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             AINSWORTH
                                                                                                                                                                                                                                                                        35
40
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 290
290
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  INNER COAT
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          J.R.,
                                                                                                                                                                                                                                                                           Mismatches
                                                                                                                                                                                                                                                                                                 Significance
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    RETROVIRIDAE;
                                                           40
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             BISHOP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 300
300
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     PROTEIN P12;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             -TFNVGW
50
                                                                                        -TFNVGW
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            PLPPSA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    310
310
                                                           50
```

Listing for Mary Hale Tue Apr 11 13:27:37 1995

58

13. SO TITIES TO BE SO OF THE STANDARD SO OF THE STANDA US-08-223-263-1 (1-353) OCH1_YEAST PROBABLE A 01-JUL-1993 01-JUL-1993 01-FEB-1994 OCHI YEAST P31755; EMBL; 01-JUL-1993 (REL. 26, CREATED)
01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
PROBABLE ALPHA-1, 6-MANNOSYLTRANSFERASE (EC 2.4.1.-). PIR; EMBL; -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ANCHORED TO THE ER AND/OR THE GOLGI MEMBRANES. -!- FUNCTION: INVOLVED IN OUTER CHAIN ELONGATION OF OLIGOSACCHARIDES OF THE TYPE MAN(9)GLCNAC(2). NAKAYAMA K.-I., NAGASU T., EMBO J. 11:2511-2519(1992) SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES. OCH1 GLYCOPROTEIN. TRANSFERASE; STRAIN=EHF-2C; SEQUENCE FROM N.A. NIAMOC TRANSMEM DOMAIN S22701; S40019; S40019. D11095; SCOCH1. PROBABLE ALPHA-1, 6-MANNOSYLTRANSFERASE 16 30 31 480 3 203 281 341 341 35155 MW; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; SIGNAL-ANCHOR; S22701 STANDARD; SHIMMA Y.-I., CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II M
LUMENAL (POTENTIAL). PRT; 480 KUROMITSU J.-R., JIGAMI Y.; A (TYPE-II MEMBRANE PROTEIN) (EC ASPARAGINE-LINKED 2.4.1.-

TLSSPAPPACDLRVLSKL--LRDS--HVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEET--KAQDI

6

70

40

Residue Gaps

20

Initial

Score Identity

0 8 8

21 20% 43

Matches = 73 Conservative Substitutions

Optimized Score

11

56 73

Significance Mismatches

7.25 236 0

CARBOHYD CARBOHYD

POTENTIAL. POTENTIAL. POTENTIAL. POTENTIAL.

1228766 CN;

CARBOHYD CARBOHYD

SEQUENCE

Page

---NIKKQI TVNKKKNQ 70 LHNLRDQLSFAFPYDSQAPIPQRVWQTWKVG-80 90 100 -ADDKNFPSSFRTYQKTW

EGX DGSDIMNWTGPG X 340

14.

US-08-223-263-1 (1-353) GAG_MSVMO POLYPROTEIN R65 (CONTAINS: CORE PROTEIN P15;

GAG MSVMO
C P03334;
C P03334;
T 21-JUL-1986 (REL. 01, CREATED)
JT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE GAG POLYPROTEIN R65 (CONTAINS: CORE PROTEIN P15; INF
GAG POLYPROTEIN P30; NUCLEOPROTEIN P10). INNER COAT PROTEIN

MOLONEY MURINE VIRIDAE; SS-RNA SARCOMA VIRUS. A ENVELOPED VIRUSES;

FIRE RANGE OF THE PROPERTY OF ONCOVIRINAE. POSITIVE-STRAND;

RETROVIRIDAE;

PROVIRUS, 82039559 SEQUENCE FROM N.A

REDDY E.P., SMITH M.J., AARONSON S.A.; SCIENCE 214:445-450(1981).

CLONE 124 CIRCULAR, SEQUENCE FROM N.A.

VAN BEVEREN C., VAN STRAATEN F., GALLESHAW J.A., VERMA I.M.; CELL 27:97-108(1981).

CEIL 27:37-108(1981).
-!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
PIR; A03932; FOMVM. POLYPROTEIN; MYRISTYLATION

COAT PROTEIN; CORE PROTEIN; CHAIN 2 131 CHAIN 132 215 CHAIN 216 478 CHAIN 479 534 ; NUCLEOPROTEIN; POLYPROTEI
CORE PROTEIN P15.
INNER COAT PROTEIN P12.
CORE SHELL PROTEIN P30.
NUCLEOPROTEIN P10.

1 1

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

-

8

Gaps Initial Score Residue Identity SO 170 180 190 200 210 220 230 TICVRRAPPTTAVPSRTSLVITLNELPNRTSGLLETNFTASARTTGSGLLK---WQQGFRAKIPGLLNQTSR 120 PPLSTPPQSSLYPAL-TPS-LGAKPKPQVLSDSGGPLIDLLTEDPPPYRDPRPP
10 130 140 150 160 170 320 330 340 350 X PTLPT-PVVQLHPLLPDPSAPTPTPTSPLLNTSYTHSQNLSQEG VGMPRDGTFNRDLITQVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPPWKKPFVHPKPPPPLLPSAPSLPLE 50 60 70 80 90 100 110 40 250 260 270 280 290 300 310 SLDQIPGY LNRI HELLNGTRGLF PGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLP LIPID VARIANT SEQUENCE MGQTVTTPLSLTL--DHWKDVERLAHN-X 10 20 519 538 n 0 0 Α, 21 24% 21 2 519 61209 Conservative Substitutions Optimized Score ¥. MYRISTATE. R -> K (IN CLONE; 1333981 CN; II --QSVDVKKRRWVTFCSAEWP--36 43 124) 8 Significance Mismatches II 7.25 114 0 Ę

15. MANB_CALSA US-08-223-263-1 (1-353) BETA-MANNANASE / ENDOGLUCANASE A PRECURSOR (MANNAN

STANDARD; PRT; 1331 B

MANB CALSA P22533; 01-AUG-1991 01-JUL-1993 01-JUN-1994

01-AUG-1991 (REL. 19, CREATED)
01-JUL-1993 (REL. 26, LAST SECUENCE UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
BETA-MANNANASE / ENDOGLUCANASE A PRECURSOR (MANNAN ENDO-1,4-BETA-MANNANASE) (ENDO-1,4-MANNANASE)
MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-MANNANASE)
ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE)).

MANA.

CALDOCELLUM SACCHAROLYTICUM. PROKARYOTA; NOT YET CLASSIFIED.

SEQUENCE FROM N.A.

APPL. GIBBS M.D., SAUL D.J., LU APPL. ENVIRON. MICROBIOL. LUTHI E., BERGQUIST P.L.;)L. 58:3864-3867(1992).

SEQUENCE 91247819 OF 1-346 FROM N.A.

LUETHI E., JASMAT N.B., GRAYLING R.A., LOVE D.R., BERGQUIST P.L.;
APPL. ENVIRON. MICROBIOL. 57:6944-700(1991).

-!- FUNCTION: DEGRADATION OF HEMICELLUIOSES, THE SECOND MOST ABUNDANT
POLYSACCHARLDES IN NATURE. CONTAINS TWO CATALITIC DOMAINS WITH
MANNANASE AND ENDOGLUCANASE ACTIVITIES.

-!- THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREE CELSIUS.

-!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
GALACTOGLUCOMANNANS.
-!- CATALYTIC ACTUVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.

Section Company 1. 18.30 S 450 70 1932 ... 4 3

Listing for Mary Hale

Tue Apr 11 13:27:38 1995 4

82

Residue Identity = Gaps = Initial Score WAQIGASNVTFKFVKLSSSVSGADYYLEIGFKSGAGQLQPGKDTGEIQMRFNKDDWSNYNQGNDWSWIQSMT 430 X 440 450 460 470 480 490 DOMAIN
ACT_SITE
ACT_SITE
CONFLICT
CONFLICT
SEQUENCE -!- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
-!- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).
EMBL; 101257; CSMANA1.
EMBL; M3063; CSMANABD.
EMBL; M37147; CSMANAB.
EMBL; M37147; CSMANAB.
ETE: A80544. EA0544. DOMAIN DOMAIN PIR; A48954; A48954.
PROSITE; PSO0659; GLYCOSYL HYDROL F5.
HYDROLASE; GLYCOSIDASE; CELLULOSE DEGRADATION; SIGNAL;
MULTIFUNCTIONAL ENZYME. CHAIN DOMAIN DOMAIN DOMAIN SIGNAL DOMAIN MELTELLLVV-1331 AA; 21 21**%** 63 41 1331 325 361 518 564 720 780 1331 162 257 Optimized Score = Matches = Conservative Substitutions 146892 MW; BETA-MANNANASE / ENDOGLUCANASE A. CATALYTIC (MANNANASE ACTIVITY). PRO/SER/THR-RICH (PT BOX). SUBSTRATE-BINDING (POTENTIAL). PRO/SER/THR-RICH (PT BOX). SUBSTRATE-BINDING (POTENTIAL). NUCLEOPHILE (BY SIMILARITY).

T -> P (IN REF. 2).

TETETET -> ROHOHRO (IN REF. 2). PRO/SER/THR-RICH (PT BOX).
CATALYTIC (ENDOGLUCANASE ACTIVITY).
PROTON DONOR (BY SIMILARITY). --MLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLHSRL 9861116 CN; 62 Significance = 85 Mismatches = ions =

--KSGAGQL--QPG---KDTG 660

670

--EIQI-RFNKSDWS

TGGQIKVLYANKETNSTTNTIRP---570 580

WIKVVN-SGSSSIDLSRVTIRYWYTVDG--ERAQSAISDWAQI

330 340 350 X
P-LLPDP----SAPTPTPT-SPLLNTSYTHSQNLSQEG

maryh@stic

stdin

NeWSprinter20

Tue Apr 11 13:29:56 1995

NeWSprint 2.5 Rev B Openwin library 3 NeWSprint interpreter 210.0

NeWSprint2.5